RL-TR-95-82
In-House Report
April 1995

WIDEBAND AUDIO COMPRESSION USING SUBBAND CODING AND ENTROPY-CONSTRAINED SCALAR QUANTIZATION

Trevor R. Trinkaus

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-95-82 has been reviewed and is approved for publication.

APPROVED: Richard S. Vonusa

RICHARD S. VONUSA, Chief
Intelligence Analysis Branch
Intelligence & Reconnaissance Directorate

FOR THE COMMANDER:

DELBERT B. ATKINSON, Colonel, USAF
Director of Intelligence & Reconnaissance

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (IRAA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1995	In-House

4. TITLE AND SUBTITLE
WIDEBAND AUDIO COMPRESSION USING SUBBAND CODING AND ENTROPY-CONSTRAINED SCALAR QUANTIZATION

5. FUNDING NUMBERS
PR - 3188
TA - 07
WU - 50

6. AUTHOR(S)
Trevor R. Trinkaus

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rome Laboratory (IRAA)
32 Hangar Road
Griffiss AFB NY 13441-4114

8. PERFORMING ORGANIZATION REPORT NUMBER
RL-TR-95-82

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (IRAA)
32 Hangar Road
Griffiss AFB NY 13441-4114

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Trevor R. Trinkaus/IRAA (315) 330-4025
Work performed at Rensselaer Polytechnic Institute under the Air Force Palace Knight Program.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Source coding of wideband audio signals for storage applications and/or transmission over band-limited channels is currently a research topic receiving considerable attention. A goal common to all systems designed for wideband audio coding is to achieve an efficient reduction in code rate, while maintaining imperceptible differences between the original and coded audio signals. In this thesis, an effective source coding scheme aimed at reducing the code rate to the entropy of the quantized audio source, while providing good subjective audio quality, is discussed. This scheme employs the technique of subband coding, where a 32-band single-sideband modulated filter bank is used to perform subband analysis and synthesis operations. Encoding and decoding of the subbands is accomplished using entropy-constrained uniform scalar quantization and subsequent arithmetic coding. A computationally efficient subband rate allocation procedure is used which relies on analytic models to describe the rate-distortion characteristics of the subband quantizers. Signal quality is maintained by incorporating masking properties of the human ear into this rate allocation procedure. Results of simulations performed on compact disc quality audio segments are provided.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Audio Compression, Source Coding, Entropy-Constrained Quantization, Subband Coding, Filter Banks	92
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to gladly thank my academic advisor, Prof. William A. Pearlman, for all his guidance and support throughout the course of this work. I would also like to express gratitude to graduate students Ulug Bayazit, Ken Melendez, and Anil Murching for their helpful suggestions and assistance with the Unix system.

Special thanks is extended to my parents for their continuous support and encouragement throughout my career. This work would certainly not be possible without them.

CHAPTER 1
INTRODUCTION AND HISTORICAL OVERVIEW

Signal compression has been the subject of extensive research for quite some time, with initial developments originating midway through the twentieth century. The work of Shannon [1] has resulted in theoretical foundations of signal compression. In his classic paper, the entropy, or information content, of a source was formulated and it was shown that the minimum transmission rate should be equal to or greater than the entropy for zero coding errors. Following this work was the eventual development of rate-distortion theory [2], [3], [4], which provided channel capacity bounds with a fidelity criterion and inspired researchers to delve further into the area of source coding research. Since that time, there has been rapid progress in the development of source coding techniques. These techniques, aimed at minimizing signal redundancy, have proven to be more efficient in terms of bit rate reduction than the more conventional techniques, such as pulse code modulation (PCM), and differential pulse code modulation (DPCM). Many of the applications have focused on the coding of video, speech, and commentary grade (7 kHz bandwidth) audio signals. The techniques receiving considerable attention, especially for the coding of speech and audio signals, were frequency domain coding techniques.

Early speech coding results due to Crochiere, Webber, and Flanagan [5], and Crochiere [6], demonstrated the advantages of partitioning the speech spectrum into bands and coding each of the bands separately using either PCM or DPCM. This technique, known as subband coding, is used in current speech and audio coding standards and is a topic of great importance in this work. Other speech coding systems made use of transform coding. These systems relied on the use of a mathmetical

transform to convert blocks of data into a representative set of transform coefficients. Zelinski and Noll [7] developed an adaptive transform coding system which took into account the changing statistics of typical speech waveforms. Speech coding results were soon complemented with the work of Johnston and Goodman [8], where a two-band subband coding system designed for transmission of bandlimited speech and music was developed. Additional improvements including real-time implementations for speech and audio coding were investigated by Cox [9] and Crochiere [10]. Although this research primarily focused on the two-band subband coding system, efforts to extend to a four-band, tree-structured system were already beginning to take shape. These subband based systems provided good subjective quality for telephone and commentary grade audio transmission at bit rates of 4 bits/sample. By the early to mid 80's, transform and other frequency domain coding systems designed for audio began to develop. At this time the revolutionary Compact Disc (CD) had arrived, which further spurred research interests in the area referred to as wideband (20 kHz bandwidth) audio coding.

The emergence of the CD marked an unprecedented achievement in the field of wideband audio reproduction. While providing very fine amplitude resolution and a large dynamic range, the CD continues to offer the highest audio quality among current audio reproduction technologies. Its 16 bit PCM format is an accepted audio representation standard [11]. The high sampling rate of CD, in addition to sample representation with a large number of bits, has made transmission of CD-quality audio difficult. At a sampling rate of 44.1 kHz, and amplitude quantization of 16 bits/sample, the resulting transmission rate of 705.6 kb/s is unacceptable for channels of limited capacity, such as mobile radio channels [11]. It is known that for CD-quality audio, the goal of sufficient data rate reduction must be accomplished by coding the data in a way such that the level of signal degradation is not perceptible

to the user. A solution to this challenging task is to hide, or *mask* distortions due to quantization, thereby making them inaudible.

The work of Brandenburg [12] resulted in an efficient frequency domain coding algorithm which operated on principles similar to transform coding systems. This algorithm included time-to-frequency conversion of an input sample block via a discrete cosine transform (DCT) and subsequent entropy coding. Auditory masking was also used to compute acceptable levels of distortion. This coding system produced transparent coding of wideband audio at 3 bits/sample. A transform coding system developed by Johnston [13] also resulted in transparent coding of wideband audio at bit rates of 4 and 3 bits/sample. This coding scheme employed a human auditory model to derive short-term spectral masking curves which were then used to extract signal redundancies. Although the exploitation of human perception characteristics had been applied earlier for the optimization of speech coding systems, the use of source models was the primary basis for efficient coding gains. Other transform coding systems and more complex multi-band subband coding systems appeared later. These systems also made use of the masking phenomena for the suppression of noise components. Good to excellent subjective audio quality was attainable at bit rates of 2.5 bits/sample and higher.

The coding methods previously mentioned have been improved upon with innovative filterbank implementations and advances in hardware that have taken place over the past few years. Such improvements along with the arrival of newer technologies and higher degrees of consumer quality expectations have opened doors to a revisiting of wideband audio coding.

In recent years, significant advances have been made in the compression of wideband audio signals. Currently, powerful algorithms are available that can

achieve efficient compression ratios while maintaining a high level of signal quality. The Moving Pictures Expert Group within the International Organization of Standardization (ISO/MPEG) has recently completed a standard for the coding of high quality digital audio. This audio coding standard is the first international standard in the field of high quality digital audio compression [11]. The MPEG audio coder is capable of providing transparent coding of CD-quality audio at bit rates of 2.67 and 2 bits/sample, or 128 and 96 kb/s at a sampling rate of 48 kHz. The need to further reduce these rates for transmission or storage applications remains an ongoing area of research.

The subject of this work involves the design and implementation of a 32-band subband coding system for CD-quality audio signals. In this system, an attempt is made to reduce the data rate to the entropy of the quantized audio source through the use of a Laplacian based rate-distortion subband model and the technique of arithmetic coding. A system block diagram can be seen in Figure 1.1. Many of the basic elements of this audio coding system are similar to those of the MPEG system. A detailed discussion of specific aspects of the system is the subject of later chapters, however a brief overview of some of the differences between this system and the MPEG system is appropriate.

The MPEG audio coding system is a subband coding system, which as stated above, falls into the class of frequency domain coding systems. A main objective of any subband coding system is the optimal or near optimal allocation of bits, or bit rate, among subbands. In the MPEG system, a dynamic rate allocation method controlled by psychoacoustic model calculations is used to determine the number of quantizer levels for a given subband. This rate allocation procedure results in the assignment of an integer number of bits to each subband. Contrary to this procedure, the rate allocation method employed in this work assigns a non-integer

number of bits. This non-integer bit assignment is interpreted as the entropy of the quantizer output of a particular subband. Its computation is based on short-term signal statistics, the subband model parameters, and masking threshold calculations. An additional difference between coders is the coding of the subband samples and appropriate side information. The MPEG coder consists of three layers of increasing complexity. In MPEG Layer I, samples are coded independently with one codeword, while in Layer II the higher frequency subbands are coded by forming groups of three samples, where each group, or "granule," is assigned a single codeword [14]. Layer III MPEG makes use of non-uniform quantization and entropy coding, where a Huffman code is used to represent the quantizer indices [14]. The distinction between the MPEG system and the system considered here comes from the use of arithmetic coding. The motivation for the use of arithmetic coding is due to greater compression efficiency than the Huffman coding technique and the possibility of reducing the data rate to the entropy of the quantized audio source. Such a reduction in rate, assuming there is no perceptible loss in signal quality, suggests a maximum compression efficiency system.

In the chapters that follow, an in-depth discussion of the components that make up the audio coding system is presented. Chapter 2 focuses on the filter bank and provides a discussion of its design and important properties. Chapter 3 discusses the concepts of subband and variable rate coding. It is in this chapter that topics such as entropy-constrained scalar quantization and arithmetic coding are presented. The use of these techniques in an adaptive audio coding scheme is outlined. In Chapter 4 an explanation of noise masking phenomena and the use of the masking threshold in subband coding is provided. Chapter 5 contains results of coding simulations implemented using audio segments taken from a compact disc. The overall performance of the system is addressed in this chapter as well. The

conclusion of the work and additional remarks are given in Chapter 6.

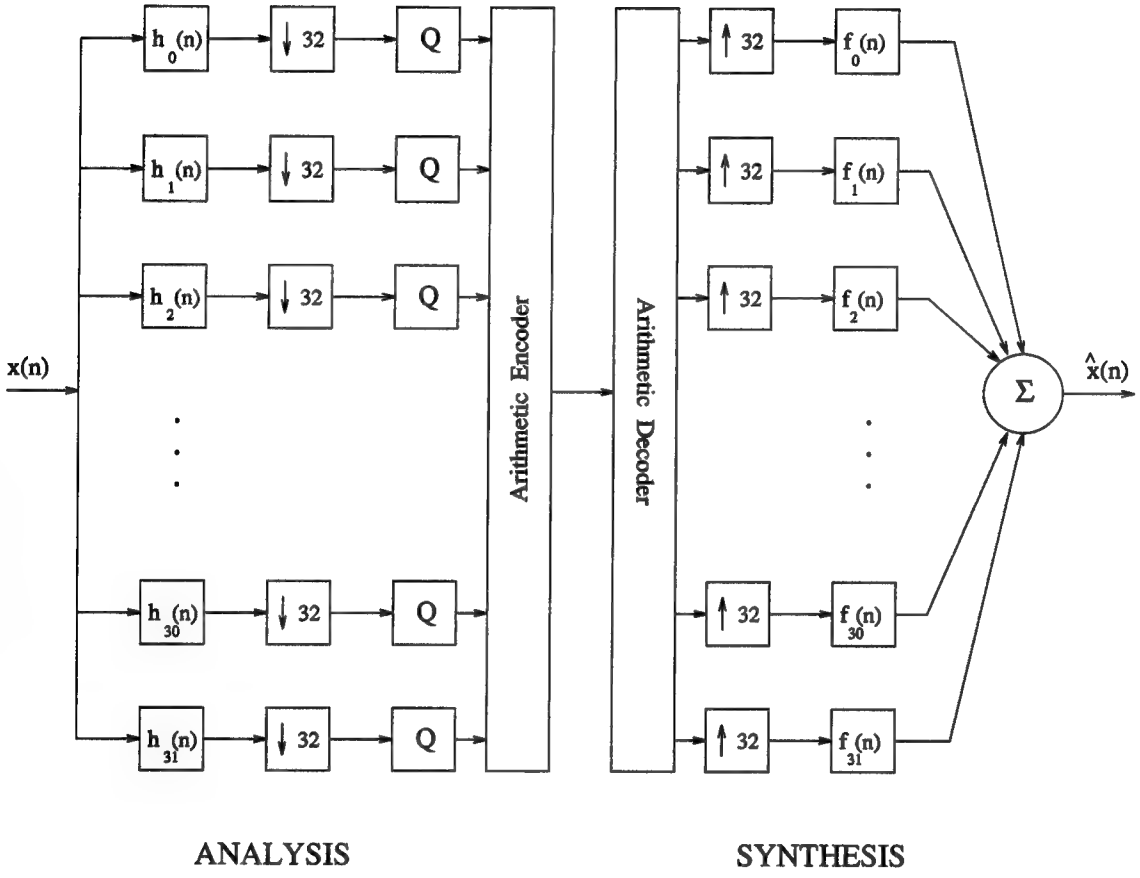

Figure 1.1: System block diagram

CHAPTER 2
SINGLE-SIDEBAND ANALYSIS-SYNTHESIS FILTER BANK

2.1 Introduction

The use of analysis-synthesis digital filter banks are of great importance in frequency domain coding systems. These systems require a frequency decomposition, or analysis of an applied input signal into contiguous frequency bands, referred to as channel signals. These channel signals, after appropriate coding and transmission, are used in the subsequent reconstruction, or synthesis of the original input. Figures 2.1 and 2.2 provide illustrations of these concepts.

Numerous advances have been made in the design and implementation of analysis-synthesis filter banks over the course of the past ten years. This research has resulted in the appearance of several filter bank structures and computationally efficient implementations, as seen in the literature [15], [16], [17], [18], [19]. The analysis-synthesis filter bank considered in this work is described as a modulated filter bank, which falls into the class of single-sideband (SSB) modulation. This particular modulation scheme results in real-valued channel signals, unlike the discrete Fourier transform (DFT) filter bank which results in complex-valued channel signals. The channel filter responses are derived from modulation of a single low-pass prototype response, a feature which requires the design of only one digital filter. Taken together, these modulated filter responses form the overall SSB analysis-synthesis filter bank. The coding system presented in this work requires the use of a filter bank consisting of 32 uniformly spaced frequency bands in the range $0 \leq \omega \leq \pi$. The following section provides an overview of the design of the filter bank from a theoretical standpoint.

2.2 Single-Sideband Filter Bank Design

The single-sideband filter bank can be derived directly from a filter bank based on a generalization of the discrete Fourier transform. This transform is known as the *generalized discrete Fourier transform* (GDFT), as stated in [15]. The mathematical description of this transform is defined in [15], and is repeated here for reference.

$$X_{GDFT}(k) = \sum_{n=0}^{K-1} x(n) W_K^{-(k+k_0)(n+n_0)},\ k = 0, 1, \ldots, K-1 \tag{2.1}$$

$$x(n) = \frac{1}{K} \sum_{k=0}^{K-1} X_{GDFT}(k) W_K^{(k+k_0)(n+n_0)},\ n = 0, 1, \ldots, K-1 \tag{2.2}$$

where

$$W_K = e^{j(2\pi/K)} \tag{2.3}$$

The quantity n_0 corresponds to the reference for the time origin, and k_0 corresponds to the reference for the discrete frequency origin. The definition of the GDFT provides a basis for the analysis-synthesis equations of the GDFT filter bank. Models of the GDFT filter bank channels can be seen in Figures 2.3 and 2.4. These channel models describe the following analysis-synthesis equations of the GDFT filter bank, defined in [15]. The analysis equation is given as

$$X_k^{GDFT}(m) = \sum_{n=-\infty}^{\infty} h(mM - n) x(n) W_K^{-(k+k_0)(n+n_0)},\ k = 0, 1, \ldots, K-1 \tag{2.4}$$

and the synthesis equation is given as

$$x_k(n) = \sum_{m=-\infty}^{\infty} f(n - mM) X_k^{GDFT}(m) W_K^{(k+k_0)(n+n_0)},\ k = 0, 1, \ldots, K-1 \tag{2.5}$$

where K represents the number of frequency channels, M is the decimation factor, and $h(n)$ and $f(n)$ are the analysis and synthesis prototype filters, respectively. The channel signals are denoted by $X_k^{GDFT}(m)$ for this particular filter bank. The channel center frequencies are located at frequencies given by the following equation.

$$\omega_k = \frac{2\pi}{K}(k + k_0), \quad k = 0, 1, \ldots, K-1 \tag{2.6}$$

These frequencies represent K equally spaced points on the unit circle in the z-domain, or equivalently the number of sample points in the GDFT. Expressions for the SSB modulated channel signals can be derived directly from the above GDFT channel signals.

As stated previously, SSB modulation results in real-valued channel signals. SSB modulation involves a frequency translation of two conjugate symmetric frequency bands from a real-valued signal, centered at $\pm\omega_k$, to the new locations $\omega = \pm\omega_{BW}/2$, where ω_{BW} is the width of the bands [15]. The resulting SSB modulated channel signal, denoted as $X_k^{SSB}(m)$, is real-valued since its spectrum remains conjugate symmetric [15]. Slight modification of the GDFT filter bank channel models results in the SSB filter bank channel models. The modifications necessary to produce the SSB channel signals are shown in Figures 2.5 and 2.6. Based on the descriptions in these figures, the equations relating the analysis-synthesis SSB channel signals to their corresponding GDFT channel signals are given as follows.

$$X_k^{SSB}(m) = Re[X_k^{GDFT}(m)e^{j\omega_{BW}mM/2}], \tag{2.7}$$

$$X_k^{GDFT}(m) = X_k^{SSB}(m)e^{-j\omega_{BW}mM/2}, \tag{2.8}$$

where $X_k^{SSB}(m)$ represents the SSB channel signal. By applying Equation 2.4 to 2.7 and Equation 2.8 to 2.5, with reference to Figures 2.5 and 2.6, the analysis-synthesis equations of the SSB filter bank result. These equations are given as

$$X_k^{SSB}(m) = Re[\sum_{n=-\infty}^{\infty} h(mM-n)x(n)e^{-j(2\pi/K)(k+k_0)(n+n_0)}e^{j\omega_{BW}mM/2}] \tag{2.9}$$

$$= \sum_{n=-\infty}^{\infty} h(mM-n)x(n)\cos(\frac{\omega_{BW}mM}{2} - \frac{2\pi}{K}(k+k_0)(n+n_0)) \tag{2.10}$$

and

$$x_k(n) = Re[\sum_{m=-\infty}^{\infty} f(n-mM)X_k^{SSB}(m)e^{-j\omega_{BW}mM/2}e^{j(2\pi/K)(k+k_0)(n+n_0)}] \tag{2.11}$$

$$= \sum_{m=-\infty}^{\infty} f(n-mM) X_k^{SSB}(m) \cos(\frac{2\pi}{K}(k+k_0)(n+n_0) - \frac{\omega_{BW} mM}{2}) \quad (2.12)$$

The design of the SSB filter bank depends primarily on the choice of k_0. This parameter determines the number of SSB channels for a particular K-point GDFT, and the bandwidths of each of the channels. It is customary to choose k_0 to be equal to a rational fraction that is less than 1. The SSB filter bank considered in this work is based on the choice of $k_0 = 1/2$. For this choice of k_0, there are $K/2$ uniformly spaced frequency bands with center frequencies located at

$$\omega_k = \frac{2\pi}{K}(k+\frac{1}{2}), \quad k = 0, 1, \ldots, \frac{K}{2} - 1 \quad (2.13)$$

and bandwidths of

$$\omega_{BW} = \frac{2\pi}{K} \quad (2.14)$$

The value of the decimation factor, M, can be determined directly from the value of K. The relationship between M and K is based on the property of critical sampling. In a critically sampled filter bank, the number of frequency domain samples is equal to the number of time domain samples. This implies that the decimation factor should be equal to the number of frequency bands, since these bands are uniform. In this case, therefore, the relationship is given as

$$M = \frac{K}{2} \quad (2.15)$$

The above equations for the analysis-synthesis SSB channel signals can now be further simplified by substituting the appropriate expressions. It will be convenient to express these equations in terms of the modulated analysis-synthesis filter responses, rather than the prototype filter responses. These equations, as defined in [15], are given as

$$X_k^{SSB}(m) = \sum_{n=-\infty}^{\infty} h(mM-n) x(n) \cos(\frac{m\pi}{2} - \frac{2\pi}{K}(k+\frac{1}{2})(n+n_0)) \quad (2.16)$$

$$= (-1)^{km} \sum_{n=-\infty}^{\infty} h_k(mM-n)x(n),\ k=0,1,\ldots,\frac{K}{2}-1 \quad (2.17)$$

and

$$x_k(n) = \sum_{m=-\infty}^{\infty} f(n-mM)X_k^{SSB}(m)\cos(\frac{2\pi}{K}(k+\frac{1}{2})(n+n_0)-\frac{m\pi}{2}) \quad (2.18)$$

$$= \sum_{m=-\infty}^{\infty} (-1)^{km} f_k(n-mM)X_k^{SSB}(m),\ k=0,1,\ldots,\frac{K}{2}-1 \quad (2.19)$$

where

$$h_k(n) = h(n)\cos[\frac{2\pi}{K}(k+\frac{1}{2})(n-n_0)] \quad (2.20)$$

and

$$f_k(n) = f(n)\cos[\frac{2\pi}{K}(k+\frac{1}{2})(n+n_0)] \quad (2.21)$$

The equations for $h_k(n)$ and $f_k(n)$ represent the modulated analysis-synthesis filter responses.

The design of the SSB filter bank is completed by substitution of appropriate values for K, M, and n_0. It was stated earlier that the filter bank consists of 32 bands. The value of M, therefore, is 32, and K, being twice this quantity, is 64. The determination of n_0 is based on orthogonality requirements. It is desirable to design the modulated filter responses in the analysis and synthesis filter banks so that they are orthogonal with each other. In other words, the dot product of each filter response with every other filter response should be equal to 0. These requirements are expressed mathematically as

$$\langle h_i(n),\ h_j(n)\rangle = 0, \quad i=0,1,\ldots,M-1, \quad j\neq i \quad (2.22)$$

and

$$\langle f_i(n),\ f_j(n)\rangle = 0, \quad i=0,1,\ldots,M-1, \quad j\neq i \quad (2.23)$$

where $h_i(n)$ and $f_i(n)$ are the modulated responses for the analysis and synthesis filter banks, respectively. The analysis and synthesis prototype filters used in this

design are also the filters used in the MPEG audio coding standard. The value of n_0 in the MPEG coder is 16. This value provides an analysis and synthesis filter bank design in which the modulated filter responses meet the above orthogonality requirements. The analysis-synthesis equations of the modulated filter responses are now given as

$$h_k(n) = h(n)\cos[\frac{2\pi}{64}(k+\frac{1}{2})(n-16)], \quad k = 0, 1, \ldots, M-1 \tag{2.24}$$

and

$$f_k(n) = f(n)\cos[\frac{2\pi}{64}(k+\frac{1}{2})(n+16)], \quad k = 0, 1, \ldots, M-1 \tag{2.25}$$

2.3 Filter Bank Properties

It is desirable in coding systems designed for compression purposes to produce a reconstructed output which is exactly equal to the input. This condition cannot be met in a practical system since signals must be quantized. In the absence of quantization, systems may be designed to produce a reconstructed output which is in fact identical to the input. This type of system, known as a perfect reconstruction (PR) system, describes the SSB filter bank used in this work. In the paragraphs that follow, a brief discussion of the prototype filter characteristics and the necessary conditions for perfect reconstruction are given.

As stated in the previous section, the analysis and synthesis prototype filters used in the design of the SSB filter bank are the same filters used in the MPEG coding system. These filters fall into the class of symmetric, finite impulse response (FIR) filters, and have tap lengths of 512. The impulse response coefficients are given in [20]. Plots of the prototype analysis impulse response and corresponding frequency response can be seen in Figures 2.7 and 2.8, respectively. Similar plots of the prototype synthesis responses are shown in Figures 2.9 and 2.10. The axes

in the frequency domain plots have been normalized with respect to the sampling frequency, or 2π. The synthesis impulse response coefficients are equal to the analysis impulse response coefficients multiplied by a factor of M. This gain is provided to compensate for attenuation introduced in the decimation processes. As seen from Figures 2.8 and 2.10, the frequency response of the prototype filters show a narrow transition band with a high side lobe attenuation exceeding 100 dB. These sharp cutoff characteristics are necessary to eliminate distortions due to aliasing and imaging. These distortions are inherent in systems involving decimation and interpolation operations. The modulated prototype frequency responses, taken together, form the analysis and synthesis filter banks shown in Figures 2.11 and 2.12, respectively. These filter banks are plotted such that every other modulated response is shown dotted for clarity reasons. The modulated responses are shifted in frequency so that there is some allowed overlap between adjacent responses. This overlap is necessary to prevent the occurance of spectral holes, or gaps, in the reconstructed output. The amount of overlap, however, must be carefully controlled so that the overall analysis-synthesis system is an identity, or perfect reconstruction, system.

A practical perfect reconstruction system is able to produce an output signal which is a delayed replica of the input signal. The expression for the output of the analysis-synthesis filter bank describes two necessary conditions which must be met in order for the filter bank to be a perfect reconstruction system. The first of these conditions, as given in [15], requires the analysis and synthesis filters to satisfy the desired input-output relation of the back-to-back filter bank. This condition is given as

$$\frac{1}{M}\sum_{k=0}^{M-1} F_k(e^{j\omega})H_k(e^{j\omega}) = 1 \quad for\ all\ \omega \tag{2.26}$$

which requires the modulated filter products to sum to 1 in the frequency domain. The terms $H_k(e^{j\omega})$ and $F_k(e^{j\omega})$ are the Fourier transforms of the modulated analysis

and synthesis responses, $h_k(n)$ and $f_k(n)$. Since there is some delay present in the output, it is sufficient for the filters to satisfy the following expression.

$$\frac{1}{M}|\sum_{k=0}^{M-1} F_k(e^{j\omega})H_k(e^{j\omega})| = 1 \quad for\ all\ \omega \tag{2.27}$$

A plot of Equation 2.27 is given in Figure 2.13, where it is shown that the modulated filters successfully satisfy this expression. This plot can be blown up, as in Figure 2.14, to reveal the presence of small oscillations in the magnitude. These oscillations, however, are insignificant as far as sufficient filter bank operation is concerned. The second requirement for perfect reconstruction was mentioned earlier, and involves the elimination of aliasing and imaging components. In most instances, this condition is automatically satisfied with proper design of the analysis and synthesis prototype filters. For situations in which $M < K$, which is the situation considered here, the cutoff frequency requirement of the prototype analysis filter is given by

$$\frac{\pi}{K} < \omega_{ch} \leq \frac{\pi}{M} \tag{2.28}$$

which ensures that aliasing will be avoided and the condition of Equation 2.27 will be satisfied. In order to avoid imaging after the interpolation processes, the cutoff frequency of the prototype synthesis filter must meet the following constraint.

$$\omega_{ch} + \omega_{cf} \leq \frac{2\pi}{M} \tag{2.29}$$

Both of the above constraints are satisfied, and the analysis-synthesis filter bank is, in fact, a perfect reconstruction system. The cutoff frequencies of both the analysis and synthesis prototype filters are equivalent, and are approximately equal to $\pi/46.5$, which lies almost halfway between the frequencies given in Equation 2.28.

The importance of the SSB analysis-synthesis filter bank and its relevance to the issue of coding will become more evident in the following chapter, where the concept of subband coding is discussed.

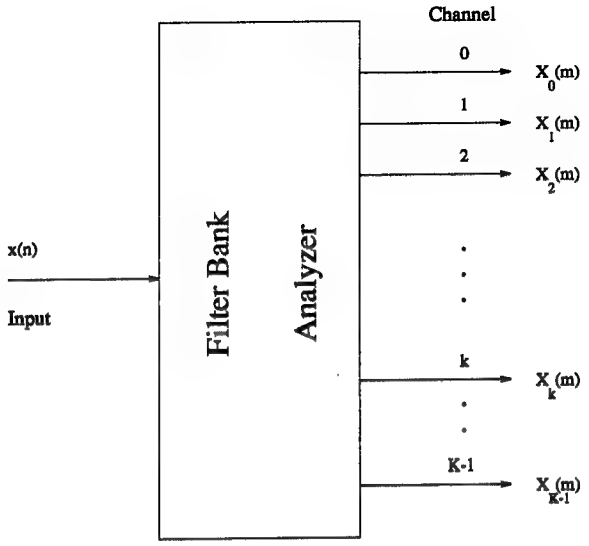

Figure 2.1: K-channel filter bank analyzer

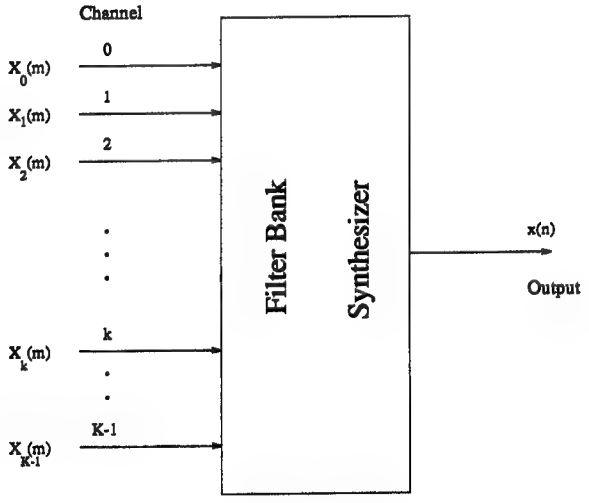

Figure 2.2: K-channel filter bank synthesizer

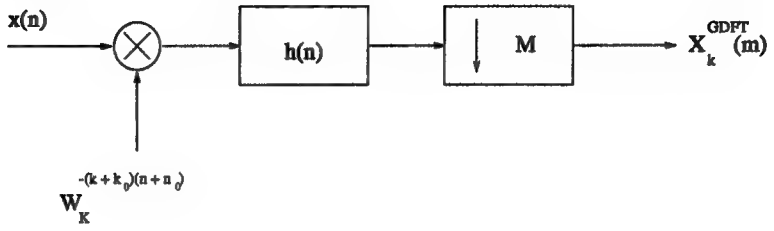

Figure 2.3: GDFT filter bank model for channel k (analyzer)

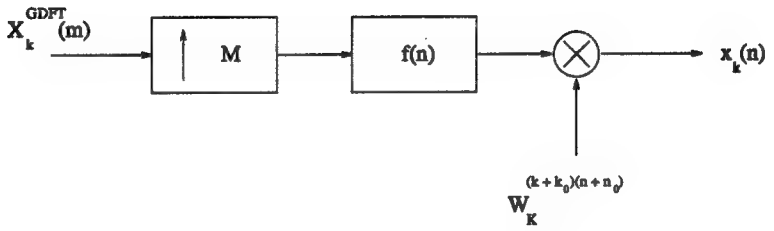

Figure 2.4: GDFT filter bank model for channel k (synthesizer)

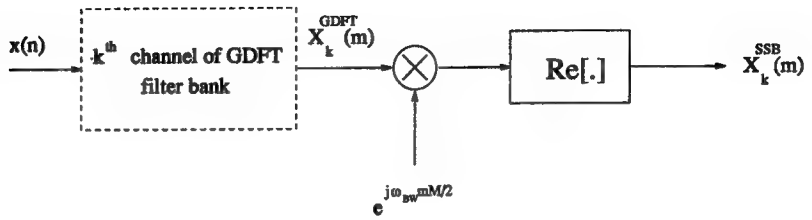

Figure 2.5: SSB filter bank model for channel k (analyzer)

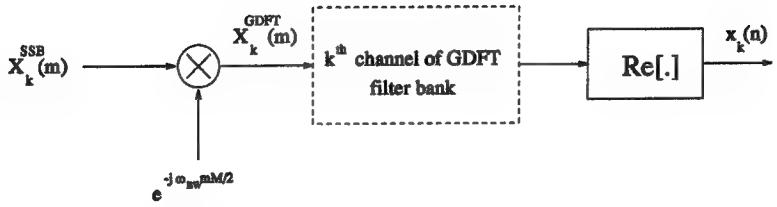

Figure 2.6: SSB filter bank model for channel k (synthesizer)

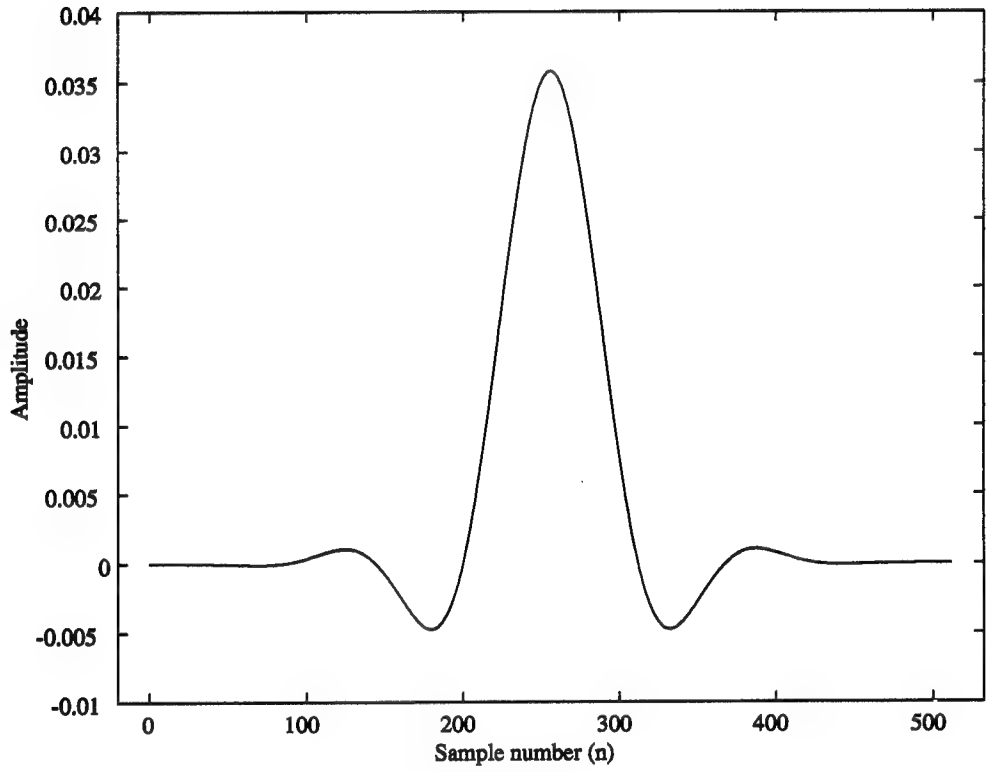

Figure 2.7: Prototype analysis filter impulse response, $h(n)$

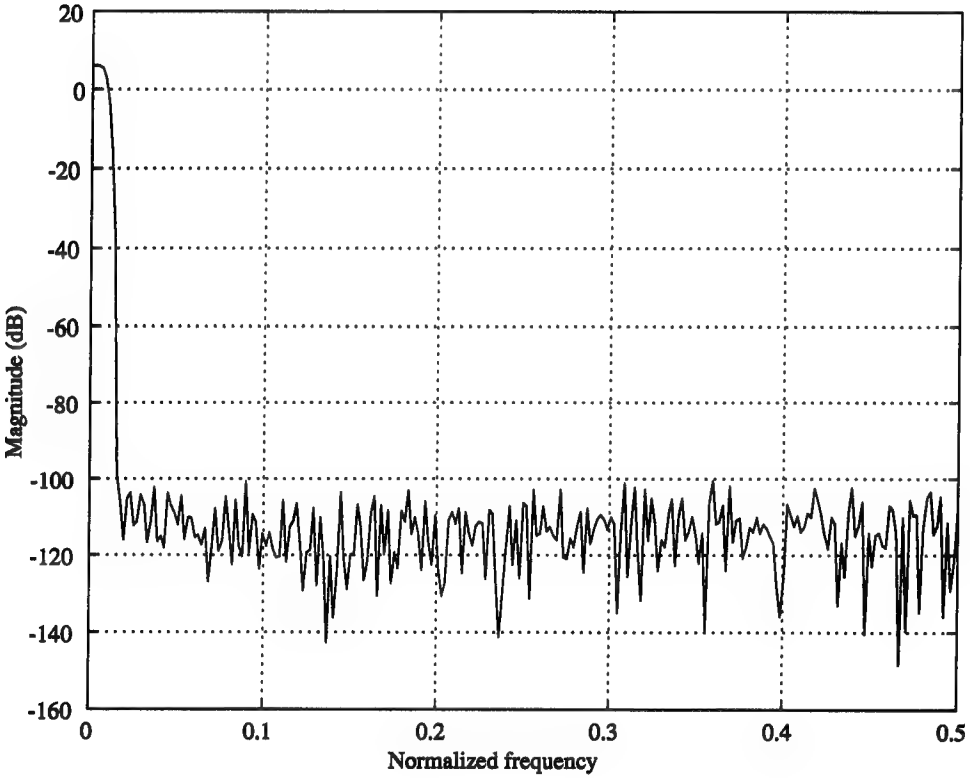

Figure 2.8: Prototype analysis filter frequency response, $H(e^{j\omega})$

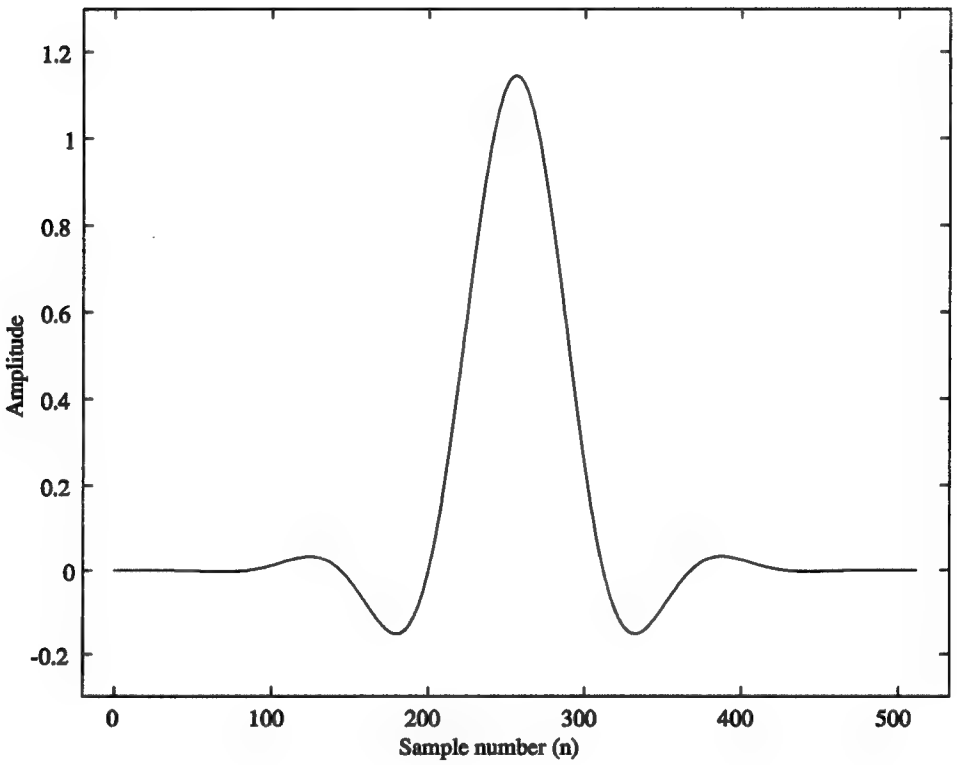

Figure 2.9: Prototype synthesis filter impulse response, $f(n)$

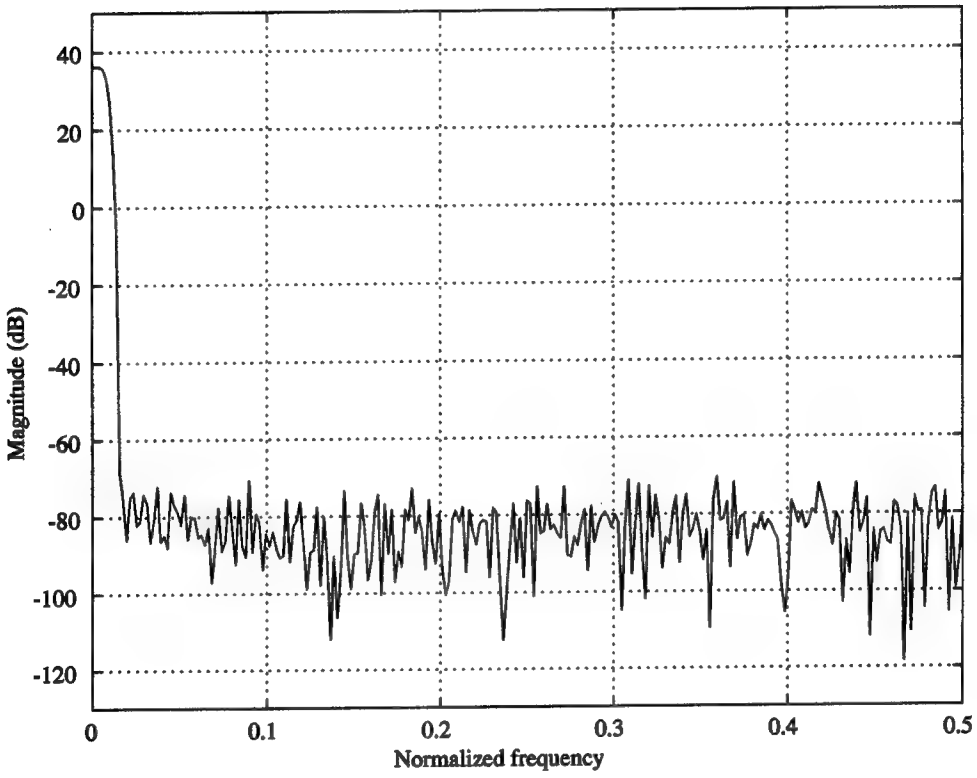

Figure 2.10: Prototype synthesis filter frequency response, $F(e^{j\omega})$

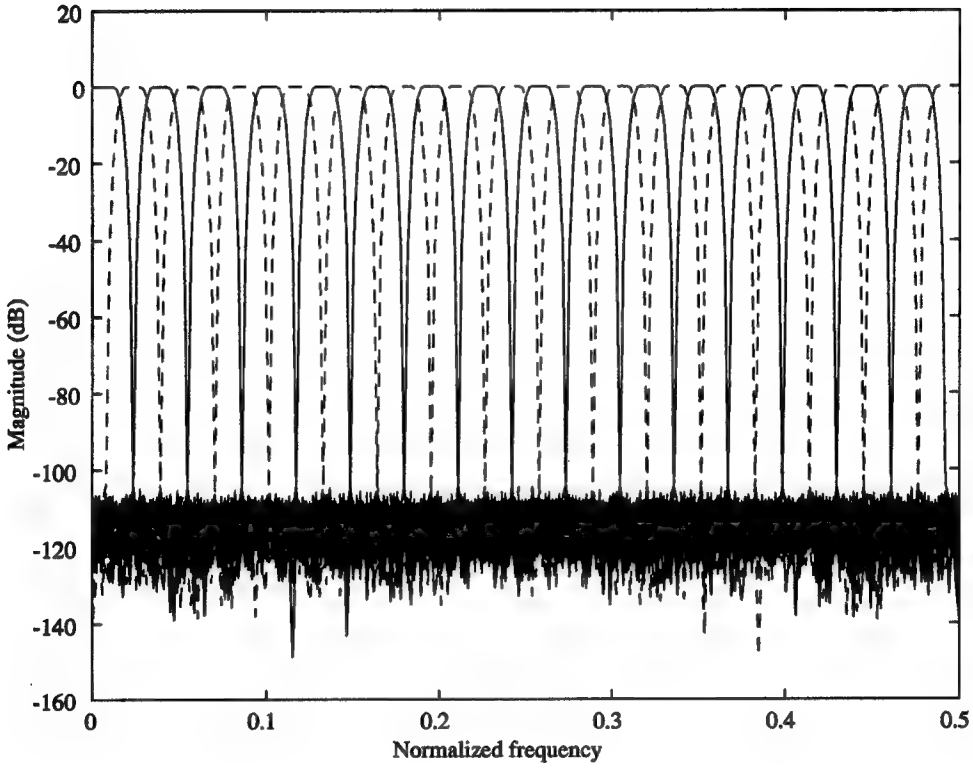

Figure 2.11: Frequency response of analysis SSB filter bank

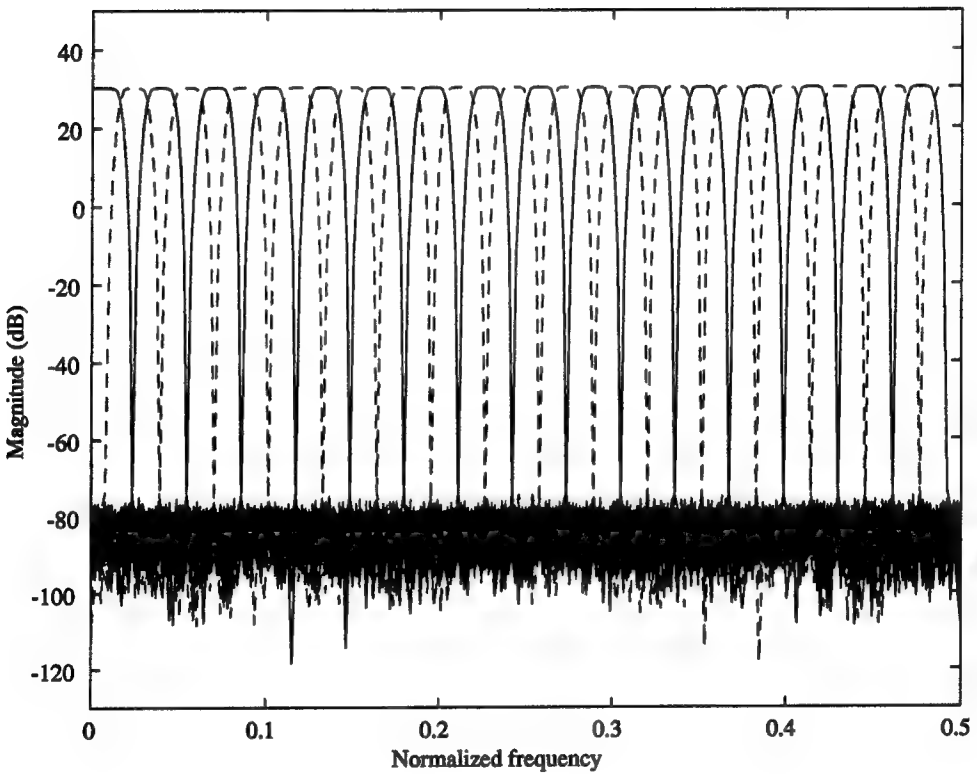

Figure 2.12: Frequency response of synthesis SSB filter bank

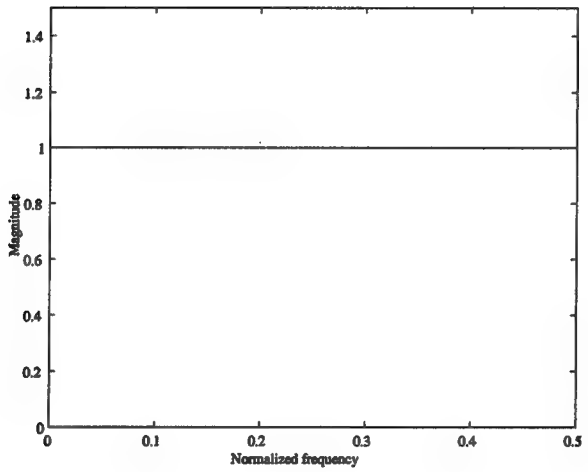

Figure 2.13: Plot of summed products of analysis and synthesis modulated frequency responses (equation 2.27)

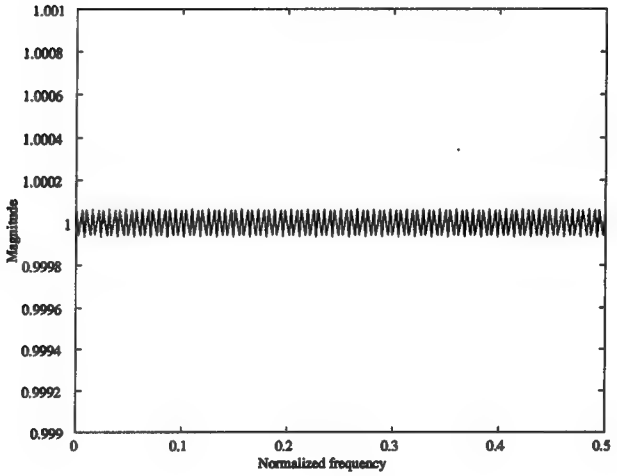

Figure 2.14: Blown up plot of equation 2.27

CHAPTER 3
VARIABLE RATE AND SUBBAND CODING

3.1 Introduction

In coding systems where it is necessary to set the bit rate to a value acceptable for bandlimited channels, the technique of variable rate coding can be applied. This coding technique is especially convenient for signals which exhibit widely varying statistical properties, such as digital audio. Since variable rate coding allows for the use of variable length binary codewords to represent quantizer output levels, it is possible to control the expenditure of bits as changes in the level of signal activity occur over time. The realization of a variable rate code can be accomplished by applying a form of noiseless source coding to a memoryless, discrete sequence, or a sequence produced by quantization. The application of noiseless coding subsequent to quantization is commonly referred to as entropy coding. The technique of entropy coding in combination with quantization, in this case scalar quantization, can achieve a bit rate which is close to the entropy of the quantized source. The performance of the quantizer is therefore measured by the entropy of its output, rather than the base two logarithm of its number of levels. It has been shown that the performance of the uniform quantizer is superior and asymptotically optimal to the performance of the non-uniform quantizer when its outputs are entropy coded. Uniform quantization, therefore, is solely the type of quantization that needs to be considered when entropy coding is used. A natural extension of these concepts leads to a more efficient form of quantization known as entropy-constrained quantization. This technique can be conveniently implemented with the use of uniform quantizers. The quantizer output entropy can be set to any desired value for a fixed number of

output levels by simply varying the probabilities of these output levels. This can easily be accomplished by varying the uniform quantizer step size. The technique of entropy-constrained quantization is precisely the technique used in the audio coding system in this work.

Entropy coding may take on various forms whose implementations differ in complexity. The particular form of entropy coding used here is known as arithmetic coding. Arithmetic coding is a technique which is suitable for the encoding of long data streams, and is capable of reducing the bit rate to the entropy of the quantized source. An important attribute of arithmetic coding is the model used to describe the probability distribution of the message to be encoded. The most efficient model is one which changes its message symbol probabilities based on the frequency of occurrence of these symbols. This type of model, known as an adaptive model, is used in the arithmetic coder in this work.

The unification of the above techniques into a system appropriate for the coding of digital audio is completed with the addition of subband coding. As stated in Chapter 1, subband coding is a particular form of frequency domain coding which acts to reduce the bit rate by minimizing signal redundancy. This is done by partitioning the spectrum of the source into frequency bands, or *subbands*, and coding each of these bands independently with a coder suitably matched to the statistics of the band. Subband coding used in conjunction with entropy-constrained quantization and arithmetic coding describes the coding system to be outlined in the following sections.

3.2 Uniform Scalar Quantization

The purpose of this section is to provide some preliminary definitions associated with scalar quantization, and to give a description of the quantizers used in

this work. Much of this information as seen in [21], is repeated here for reference.

Scalar quantization is a mapping of samples from a memoryless source, X, with probability density function, $f_X(x)$, to some reproduction value, y_k, in a finite set of reproduction values, $\mathcal{C} = \{y_1, y_2, \ldots, y_N\}$. The set $\mathcal{C}$ is otherwise known as the codebook. This mapping is subject to a distortion criterion which basically governs the choice of reproduction value, y_k, for a given source sample x. The distortion criterion, or measure, considered here is the mean-squared error criterion, given by

$$d(x, y_k) = (x - y_k)^2 \tag{3.1}$$

The mapping of source samples is based on the principle of minimum distortion, which states that a given source sample, x, will be mapped to reproduction value, y_k, if $d(x, y_k) \leq d(x, y_l)$ for all $l \neq k$. It is convenient to define this mapping in terms of intervals, or bins, as given in [21]. These intervals are specified as

$$I_k = \{x \ : \ d(x, y_k) \leq d(x, y_l), \text{ all } l \neq k\}, \qquad k = 1, 2, \ldots, N \tag{3.2}$$

which is a partition representing the set of $x's$ that are nearest in distortion measure to y_k for each y_k in $\mathcal{C}$. The partition can also be given in terms of its endpoints as

$$I_k = [x_{k-1}, x_k) = \{x \ : \ x_{k-1} \leq x < x_k\}, \qquad k = 1, 2, \ldots, N \tag{3.3}$$

which may be used to provide a more compact definition of the quantization process. This definition, as seen in [21], is given as

$$y_k = Q(x) \quad \text{if } x \in I_k \tag{3.4}$$

The reproduction values $\{y_1, y_2, \ldots, y_N\}$ will now be referred to as quantizer levels, and the interval endpoints $\{x_0, x_1, \ldots, x_{N-1}, x_N\}$ will be referred to as decision thresholds [21]. The quantization error is determined by computing the average of

$d(x, Q(x))$. This error is more commonly referred to as the distortion, denoted by D, and is given as

$$D = E[d(x, Q(x))] = \int_{-\infty}^{\infty} d(x, Q(x)) f_X(x) dx \tag{3.5}$$

Applying equation 3.4 to the above results in

$$D = \sum_{k=1}^{N} \int_{x_{k-1}}^{x_k} d(x, y_k) f_X(x) dx \tag{3.6}$$

where $d(x, y_k)$ is given by Equation 3.1, and $f_X(x)$ is the source probability density function (pdf). The rate of the quantizer, measured in bits/sample, is defined as

$$R = \log_2(N) \tag{3.7}$$

which gives the number of binary digits needed to represent one of N quantizer levels.

The process of uniform scalar quantization may involve uniformly spaced decision thresholds, quantizer levels, or both. The uniform quantizers used in this work consist of uniformly spaced decision thresholds and quantizer levels. These quantizers are mid-tread, meaning there is a quantizer level, y_k, located at 0. The number of quantizer levels, N, is odd in this case and is taken to be 255. The quantizer levels are specified as

$$y_k = a + (k-1)\Delta, \qquad k = 1, 2, \ldots, N \tag{3.8}$$

The decision thresholds are placed at the midpoints of the intervals specified by consecutive quantizer levels. They are given as

$$x_k = a + (k - \frac{1}{2})\Delta, \qquad k = 1, 2, \ldots, N-1 \tag{3.9}$$

Since the quantizers used are of the mid-tread type with 255 levels, a is given as

$$a = -\frac{N-1}{2}\Delta = -127\Delta \tag{3.10}$$

The quantizer step size, Δ, is chosen to minimize D for a fixed number of levels, N. The topic of the following section describes a technique in which the step size is chosen to minimize the distortion while the entropy of the quantizer output remains fixed. This technique is known as entropy-constrained quantization.

3.3 Entropy-Constrained Scalar Quantization

Entropy-constrained scalar quantization (ECSQ) is suitable for systems which make use of some form of entropy coding subsequent to scalar quantization. When entropy coding is used, it is no longer meaningful to evaluate the performance of the quantizer using Equation 3.7 as a measure of rate. Also since entropy coding is noiseless, the only error that results is due to quantization. The entropy, therefore, in addition to the distortion, are the quantities used to evaluate quantizer performance.

The quantizer output may be regarded as a discrete amplitude, memoryless source. Each quantizer level has a probability associated with it which can be computed analytically when the source pdf is known. These probabilities are given as

$$P(y_k) = \Pr\{x_{k-1} \leq X < x_k\} = \int_{x_{k-1}}^{x_k} f_X(x)dx, \qquad k = 1, 2, \ldots, N \tag{3.11}$$

The entropy of this source, measured in bits/sample, is defined as

$$H = \sum_{k=1}^{N} P(y_k) \log_2(\frac{1}{P(y_k)}) \tag{3.12}$$

$$= -\sum_{k=1}^{N} P(y_k) \log_2(P(y_k)) \tag{3.13}$$

The entropy will always be less than the rate defined in Equation 3.7 since the quantizer level probabilities are not equal. These probabilities can be allowed to vary by simply varying the interval spacing in Equation 3.11 above. For the uniform quantizer, this interval spacing is the step size, Δ. Variations in quantizer level

probabilities will reflect changes in the entropy of the quantizer output. Therefore, for a fixed number of quantizer levels, the distortion may be constrained to some limit, while the step size is adjusted to minimize the entropy. Similarly, the entropy may be constrained not to exceed some value, while the quantizer step size is adjusted to minimize the distortion. For every fixed limit on the quantizer entropy, there exists a step size, for some minimum value of N, which corresponds to a minimum distortion value. The same distortion value can be achieved for larger values of N for the same entropy, as long as that entropy is less than $\log_2(N)$. In practice, N is chosen to be sufficiently large to accomodate the desired range of entropies. As the limit on the entropy is allowed to vary over a range, the minimum distortion and corresponding step size may be determined over this range as well. Since the decision thresholds and quantizer levels are both uniformly spaced, a simple solution to this task is to vary the quantizer step size and calculate the entropy and distortion values which correspond. For a given source pdf, an entropy-distortion characteristic can be determined using this technique. For each quantizer step size, Equations 3.13 and 3.6 may be used to compute corresponding entropy and distortion values. Therefore, it can be determined beforehand which step size is needed to achieve minimum distortion for a fixed entropy. This method was developed by Goblick and Holsinger [22] who showed that quantizers with uniformly spaced output levels are nearly optimum for the quantization of Gaussian sequences using the squared error criterion. It was later shown by Gish and Pierce [23] that, for large rates, uniform quantization with the output levels located at the midpoints of the quantization intervals resulted in optimal performance for any density function and squared error criterion. In this work, a technique is employed in which each subband signal is quantized using a uniform quantizer whose step size is determined through the assignment of subband rates. These rates correspond to predetermined values of

the entropy of the quantizer output for the uniform quantization of a unit variance Laplacian source. The choice of the Laplacian density may be attributed to Berger [24], who showed that quantizers of a fixed entropy rate with uniformly spaced thresholds are truly optimum under the squared error criterion for data having either an exponential or Laplacian distribution. The representation of the quantizer levels by the midpoints of the quantization intervals is a convenient approximation and does not significantly degrade performance unless the rate of the quantizer is small. The Laplacian pdf is given as

$$f_X(x) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}}{\sigma}|x|} \tag{3.14}$$

where σ is the standard deviation. The entropy-distortion characteristic for the unit variance ($\sigma^2 = 1$) Laplacian, using the squared error distortion measure and taking N to be 255, can be seen in Figure 3.1. This characteristic is also shown in Figure 3.2, where the distortion is expressed in units of decibels. Since the quantizers used are symmetric, the use of an odd number of levels allows obtainable entropies below 1 bit/sample. Taking N to be equal to 255 levels results in a range of rates that is both practical and efficient, since each quantizer index may be represented using 8 bits. The application of entropy-constrained scalar quantization to subband coding will be discussed in the closing section of this chapter. Following, is a discussion of arithmetic coding, a form of entropy coding used in the subsequent encoding of the quantizer output.

3.4 Arithmetic Coding

Arithmetic coding is a noiseless compression technique in which a code string representing a fractional value between 0 and 1 is used to depict the encoded data [25]. Successive data symbols are encoded according to a probability model used to

describe the frequency of occurrence of each symbol. Compression is achieved by assigning shorter codewords to the more probable symbols, and longer codewords to the less probable symbols, similar to the Huffman coding technique. In arithmetic coding, however, the symbol probabilities are not restricted to be integral powers of 1/2 in order to achieve optimum performance, as in Huffman coding. The efficiency of the arithmetic code depends on the accuracy of the model used to represent the data. The most efficient model is one which adapts to the changing symbol statistics by updating itself as new data symbols appear over successive iterations. A fixed model can also be used, however a reduced amount of compression usually results. There are some tradeoffs between the uses of these models. While the adaptive model may result in greater compression, its implementation is a bit more complex, which tends to reduce the speed of operation of the coding algorithms. The fixed model is more robust and easier to implement since it is unnecessary to update the model as new data symbols are encoded. The use of the fixed model speeds up algorithm operation, but results in less compression.

The algorithms used to peform the encoding, or decoding, operate on one data symbol per iteration. The encoding process is accomplished by performing successive subdivisions of the unit interval into regions which correspond to the individual symbol probabilities. Each encoding iteration consists of inspecting a new data symbol, determining its probability, and subdividing the current interval based on the value of the new symbol's probability. In this way, the entire data stream can be represented by a code string which is equivalent to a real fraction between 0 and 1. As the number of data symbols in the stream increases, the interval needed to represent them becomes smaller. It is, therefore, necessary to modify the algorithm to use fixed precision arithmetic. The decoding is accomplished by undoing the operations of the encoder once the final interval which represents the entire code

string is known. This is carried out by performing magnitude comparisons between the code string and the intervals allocated for the data symbols during each decoding iteration. The intervals are known to the encoder and decoder, and are based on the symbol probabilities. The magnitude of the code string indicates the width of the interval, which allows the decoder to determine which symbol was sent upon inspection of this interval during each iteration.

The arithmetic coding implementation in this work uses an adaptive source model. Details of the implementation, along with a tutorial overview, are given in [26].

3.5 Subband Coding

The previous quantization and coding schemes are elegantly linked into a coding system suitable for digital audio with the addition of subband coding. As previously stated, subband coding is a frequency domain coding technique in which an input signal is decomposed into spectral components, and each of these components is coded separately. The spectral decomposition is accomplished using an analysis filter bank, as discussed in Chapter 2. Following the decomposition is a downsampling operation which causes the filter bank output sequences to become full band sequences at a lower sampling frequency. The combination of filtering and downsampling is known as decimation. The decimated outputs are referred to as subband signals. The subband signals are then suitably coded and are passed through a synthesis filter bank, subsequent to a decoding operation. The filtered subband outputs are then upsampled, which increases the sampling frequency of the subband signals to that of the input signal. The process of filtering and upsampling is described as interpolation. Reconstruction of the original input is accomplished by summing the interpolated outputs.

The coding of each subband component depends greatly on its spectral content, which is commonly assessed using its variance. Since these subband variances are different for every subband, the bit rates required to code each of the subbands vary as well. Although it is common to simply vary the number of quantizer levels in order to meet the required bit rate of each subband, this method imposes the restriction that the bit rate be an integer value. This restriction may be overcome by using entropy-constrained quantization for a fixed level quantizer, as discussed above. This method is well suited for subband coding and does not require the subband rates to be integer values. Once the required rate for each subband is determined, a uniform quantizer is selected whose entropy matches this rate. This is equivalent to choosing a quantizer step size from a predetermined table of step sizes which gives the required entropy. The determination of subband rates is accomplished using a rate allocation algorithm which is based on analytic expressions for the rate-distortion characteristics of the subbands. This algorithm, as seen in [21], is described in the following section.

3.5.1 Rate Allocation to Subbands

In a subband coding system, it is desirable to distribute rates among subbands such that the overall distortion takes on its minimum value for a desired overall rate. Since it is assumed that the subbands are uncorrelated, the relationship between the subband rates and the overall rate of the system is given as

$$R = \frac{1}{M} \sum_{m=0}^{M-1} r_m \tag{3.15}$$

where r_m is the rate of the m^{th} subband, M is the number of subbands, and R is the code rate of the system measured in bits/sample. A similar expression for the distortion may be derived. This expression is given in terms of the subband distortions which are based on the squared-error criterion. These terms will otherwise

be referred to as the subband's mean-squared error. The term $d_m(r_m)$ will be used to represent the mean-squared error of a unit variance subband whose rate is r_m. The actual subband mean-squared error may be obtained by simply scaling $d_m(r_m)$ by the subband variance, σ_m^2. The reduction of per-sample distortion after interpolation is compensated by the gain of the synthesis filter bank to give an overall distortion after synthesis of

$$D = \sum_{m=0}^{M-1} \sigma_m^2 d_m(r_m) \tag{3.16}$$

where D is the average distortion per sample.

There are useful algorithms for allocating rate among subbands whose methods do not rely on assumptions of the subband rate-distortion characteristics [27], [28]. In this system, however, the allocation of rate among subbands is performed by using an analytic expression to model the rate-distortion characteristic of each subband. The following equation, as seen in [21], represents a model which is acceptable for describing the rate versus mean-squared error characteristic for the scalar quantization of a unit variance signal.

$$\rho(r) = g(r)2^{-ar}, \quad r \geq 0 \tag{3.17}$$

where $g(r)$ is defined such that $g(0) = 1$, and a is a constant whose value cannot exceed 2. It is convenient to regard $g(r)$ as a constant term, since it is usually much more slowly varying than the exponential 2^{-ar}. This term will be denoted by g. Equation 3.17 is now given as

$$\rho(r) = g\,2^{-ar}, \quad r \geq 0 \tag{3.18}$$

where the constants g and a are determined from a linear fit to the natural logarithm of the experimentally determined curve in Figure 3.1. Results of the curve fit give

the following values for g and a

$$\begin{aligned} g &= \exp(0.22067) \approx 1.247 \\ a &= 1.38382/\ln(2) \approx 1.996 \end{aligned} \tag{3.19}$$

Figure 3.3 shows a plot of Equation 3.18 using the above parameters, along with the plot shown in Figure 3.1. Figure 3.4 shows the same plots on a log scale. It can be seen from these figures that the model fits the data fairly well, especially for rates above 0.5 bits/sample.

The model of Equation 3.18 with the parameters given in (3.19) may now be used to define the unit variance subband mean-squared error term, $d_m(r_m)$, given in Equation 3.16. This relationship is expressed as

$$d_m(r_m) = \rho(r_m) = g\, 2^{-ar_m} \tag{3.20}$$

The average distortion of Equation 3.16 may now be rewritten as

$$D = g \sum_{m=0}^{M-1} \sigma_m^2 2^{-ar_m} \tag{3.21}$$

In order to determine the optimal allocation of rates, it is necessary to determine the set of rates which result in the lowest average distortion, D, for a given code rate, R. This problem may be solved analytically through the use of the Kuhn-Tucker Theorem [21], which states the necessary and sufficient conditions for a minimum point of a convex upward function defined over a convex upward space. The solution, which is derived in [21], is given as follows

$$\begin{aligned} M\sigma_m^2\rho'(0) &\leq \frac{S}{R}, \quad r_m = 0 \\ M\sigma_m^2\rho'(r_m) &= \frac{S}{R}, \quad r_m > 0 \end{aligned} \tag{3.22}$$

where S/R is interpreted as the scaled slope of a subband rate-distortion curve evaluated at a particular rate, r_m. The prime is used to indicate a derivative with

respect to the argument. This solution may be restated as

$$\begin{aligned} (\sigma_m^2 g/n_m)2^{-ar_m} &= \theta \quad \text{for all } r_m > 0 \\ \sigma_m^2 g/n_m &\leq \theta \quad \text{for all } r_m = 0 \end{aligned} \tag{3.23}$$

where n_m is the number of samples per subband and $\theta \equiv -S/a\ln(2)Mn_mR$. Solving the above for r_m results in

$$r_m = \begin{cases} \frac{1}{a}\log_2[(\sigma_m^2 g/n_m)/\theta] & , \ \sigma_m^2 g/n_m > \theta \\ 0 & , \ \sigma_m^2 g/n_m \leq \theta \end{cases} \tag{3.24}$$

which describes the rate allocation using the model defined in (3.18), and the parameters defined in (3.19). The parameter, g, drops out by redefining θ as

$$\theta' = \frac{\theta}{g} \tag{3.25}$$

Equation 3.24 above may, therefore, be more compactly written as

$$r_m = \begin{cases} \frac{1}{a}\log_2[(\sigma_m^2/n_m)/\theta'] & , \ \sigma_m^2/n_m > \theta' \\ 0 & , \ \sigma_m^2/n_m \leq \theta' \end{cases} \tag{3.26}$$

The parameter θ' can be determined iteratively by choosing an initial distortion range and progressively narrowing this range as the rate, R, falls above or below the desired rate. This technique is known to converge to the desired (R, D) point rather quickly.

The subject of the following chapter involves the application of noise masking phenomena to the rate allocation rule defined in (3.26). A description of the masking threshold will be given, and it will be shown how this threshold may be used as a frequency weighting of the noise in each subband.

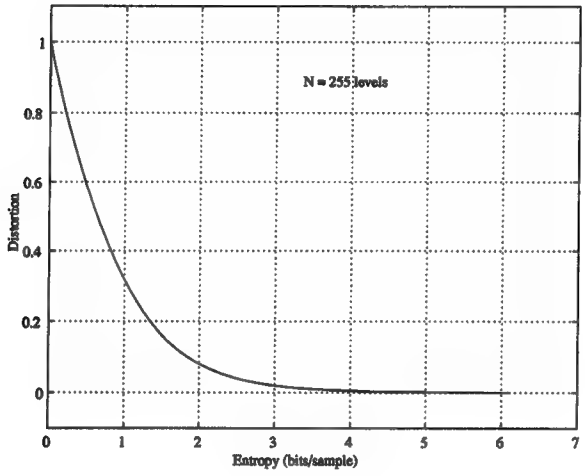

Figure 3.1: Distortion versus entropy characteristic for the scalar quantization of a unit variance Laplacian signal

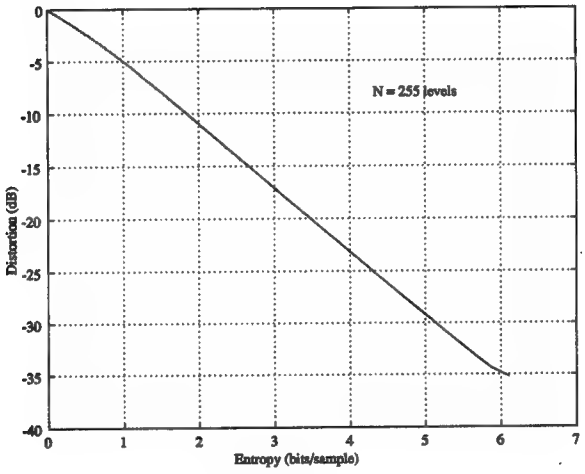

Figure 3.2: Distortion versus entropy characteristic of Figure 3.1 on a log scale

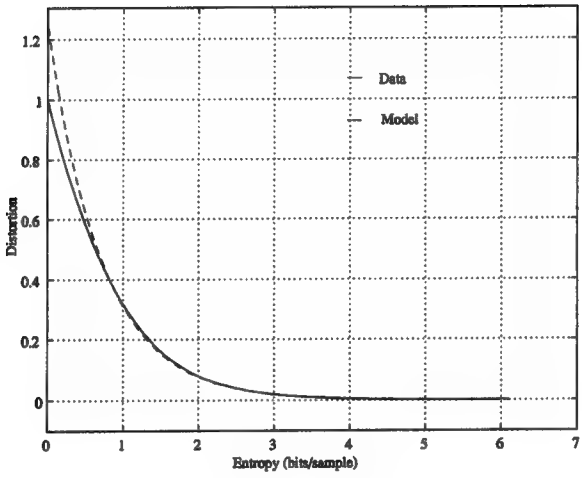

Figure 3.3: Comparison of the unit variance rate-distortion model with the curve of Figure 3.1

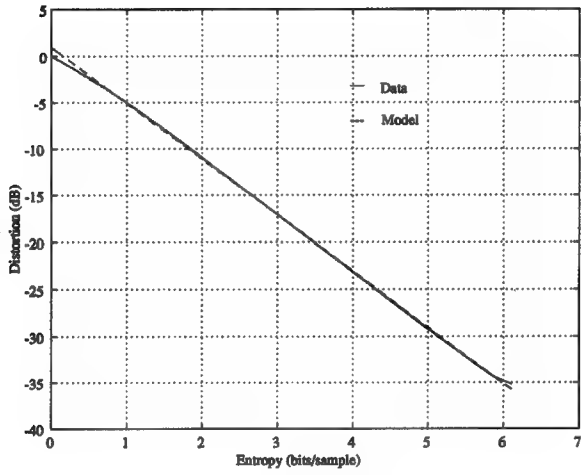

Figure 3.4: Comparison of the logarithm of the unit variance rate-distortion model with the curve of Figure 3.2

CHAPTER 4
PSYCHOACOUSTIC MASKING

In almost all audio coding systems, properties of human perception are exploited in some way. This is often accomplished by employing a coding method which relies on a special form of masking, known as simultaneous masking. Simultaneous masking is the phenomenon that a weak signal is made inaudible, or masked, by a simultaneously occurring stronger signal. In audio coding the weak signal may represent quantization noise or aliasing distortion, and the stronger signal is often a complex tone. The signal to be masked is referred to as the target, and the masking signal is referred to as the masker [29]. Masking occurs when the level of the target falls below what is known as the masking threshold.

4.1 Masking Threshold

The masking threshold is derived from a threshold of hearing in the absence of a masker, which is known as the threshold in quiet. The threshold in quiet describes as a function of frequency the level of a pure tone that is just audible [30]. The term level refers to sound pressure level (SPL), which is measured in decibels. The threshold in quiet plotted over a logarithmic frequency axis is shown in Figure 4.1. When a noise target occurs in the presence of a masker, the threshold of audibility of the target is raised over the threshold in quiet for frequencies near the frequency of the masker. This raised threshold is called the masking threshold. Targets whose sound pressure levels lie below the masking threshold are masked. In audio coding, targets are usually noise sources due to quantization. The goal for systems designed for high quality coding is to try to keep these noise sources below

the masking threshold so that they will not be audible. As a simple illustration of the masking effect, the masker is considered to be a pure tone. Figure 4.2 shows a plot of the masking threshold for a pure tone masker at 3.5 kHz. The dotted curve in this figure is the threshold in quiet. It can be seen from this plot that the slope of the masking threshold is steeper for lower frequencies, which exemplifies the fact that higher frequencies are more easily masked.

A more practical masking threshold will appear quite different from the one shown in Figure 4.2. Source signals often contain multiple maskers and targets, and the maskers are not always described as tonal. Audio signals consist of many maskers whose type may be tonal or noise-like. Figure 4.3 shows a typical masking threshold for a fragment of classical music. The lower curve in Figure 4.3 is the threshold in quiet. The length of the audio fragment used to compute this threshold was 23.2 ms. Figure 4.3 demonstrates a broad masking effect which is characterized by frequent peaks and dips. This type of effect is typical for signals such as audio, since tonal components generally consist of many harmonics which often occur at the same time.

Exploitation of the masking effect involves confining the coding errors to lie below the masking threshold, as previously stated. Since the human ear distinguishes sounds over limited frequency bands, called critical bands, it is only logical to attempt to mask noise targets which occur within these bands. As audio content and corresponding masking thresholds vary over frequency, so do critical band noise levels. It is, therefore, necessary to have control over the accuracy with which certain frequency components are coded, as in subband coding. The masking of noise targets in subband coding can be best explained by introducing the quantity known as signal-to-mask ratio (SMR). Signal-to-mask ratio is defined as the ratio

of the signal power to the masking threshold, or as the difference of the corresponding levels in decibels [11]. Within a particular critical band, noise targets will be masked, or made inaudible, if the signal-to-noise ratio (SNR) in that band exceeds the SMR. An illustration of this is provided in Figure 4.4. Also shown in this figure is the quantity called noise-to-mask ratio (NMR), which, if positive, measures the perceivable distortion in a given critical band. In subband coding, it is desirable to mask noise targets within subbands. This can be accomplished by determining the SNR required to just exceed the SMR in a given subband. Determination of the subband SMR, however, depends directly on the masking threshold.

In order to accurately compute the masking threshold, it is necessary to estimate the audio signal's short-time power spectral density (PSD). Therefore, the masking threshold is typically computed every $10 - 30$ ms [29]. In this work, the masking threshold is computed using the procedure described in [20]. This procedure may be briefly summarized as follows. First, the short-time PSD is estimated using a Hann window and a 1024-point FFT. Based on an analysis of the PSD, the maskers present in the signal are identified. Next, the individual masking thresholds of the critical bands are determined. These thresholds depend on the type and sound pressure level of the masker, and also the frequency range of each of the critical bands. The total masking threshold is then computed by adding the individual masking thresholds of the critical bands and the threshold in quiet. The subband SMR can be computed by taking the difference between the maximum signal sound pressure level and the minimum masking threshold sound pressure level within each subband.

As explained in the previous chapter, the calculation and assignment of subband rates are based on a subband rate-distortion model. Therefore, it is inappropriate to simply assign rates to subbands which provide an SNR that just exceeds

the SMR, as given by the masking condition above. Alternatively, the subband SMR values are used to weight the noise within each subband in an attempt to simultaneously satisfy the rate constraint and the masking condition. The previous description of the rate allocation given in Equation 3.26 is, therefore, in need of modification as explained in the following section.

4.2 Frequency Weighting of Noise in Subband Coding

One of the benefits of subband coding is the ability to shape the spectrum of inherent coding noise. This is done by assigning rates which reflect the amount of distortion associated with a particular subband quantizer. Subband rate assignment can be influenced by appropriately weighting the subband quantization distortion, or in this case, the mean-squared error. The subband signal-to-mask ratio provides a reasonably good indication of the accuracy with which a particular subband should be coded. The SMR is high when the masking threshold is low, which indicates that distortions due to coding errors are close to becoming audible. On the contrary, the SMR is low when the masking threshold is high, which indicates that distortions are more likely to lie below the masking threshold, and therefore, remain inaudible. The signal-to-mask ratio can be used to place more emphasis on subbands which are in need of accurate coding to achieve a high signal-to-noise ratio, and less emphasis on subbands where additional coding accuracy may not be useful. The unit variance subband mean-squared error is given in (3.20), and is defined as

$$d_m(r_m) = g\, 2^{-ar_m} \tag{4.1}$$

where g and a are given in (3.19). This equation must be slightly modified when subband noise weighting is used. The subband weights will be denoted by w_m and are equal to the subband signal-to-mask ratios on a linear, rather than dB, scale.

According to the masking condition the subband SNR must exceed the subband SMR, which may be stated as follows

$$\begin{aligned} SNR_m &> SMR_m \\ 10\log_{10}\left(\frac{\sigma_m^2}{\sigma_m^2 d_m(r_m)}\right) &> 10\log_{10}(w_m) \\ \frac{1}{d_m(r_m)} &> w_m \end{aligned} \tag{4.2}$$

where the term, $\sigma_m^2 d_m(r_m)$, represents the subband mean-squared error, and the subband weights, w_m, are given as

$$w_m = 10^{SMR_m/10} \tag{4.3}$$

Since the subband weights are simply used to scale the subband mean-squared error, the expression in (4.1) for the subband mean-squared error becomes

$$w_m \sigma_m^2 d_m(r_m) = w_m \sigma_m^2 g\, 2^{-ar_m} \tag{4.4}$$

and the previous solution to the Kuhn-Tucker Theorem is now given as

$$\begin{aligned} (w_m\sigma_m^2/n_m)2^{-ar_m} &= \theta' \quad \text{for all } r_m > 0 \\ w_m\sigma_m^2/n_m &\leq \theta' \quad \text{for all } r_m = 0 \end{aligned} \tag{4.5}$$

where θ' is defined in (3.25). The new rate allocation, therefore, is

$$r_m = \begin{cases} \frac{1}{a}\log_2[(w_m\sigma_m^2/n_m)/\theta'] & , \quad w_m\sigma_m^2/n_m > \theta' \\ 0 & , \quad w_m\sigma_m^2/n_m \leq \theta' \end{cases} \tag{4.6}$$

which is the same as the previous rate allocation given in (3.26), with the exception of the subband weighting terms. According to the masking condition given in (4.2), the product $w_m d_m(r_m)$ must be less than 1 for masking to occur. Simulations have shown that, on average, this condition is satisfied for the lowest 1/3 of the audio frequency range when the code rate, R, is set to 2 bits/sample. This range consists of approximately the lowest ten subbands, and covers the frequencies 0−6890 Hz. Since

most higher frequency subbands contain less power, lower rates are often assigned to them, which results in a larger mean-squared error. The masking condition, therefore, is usually not satisfied for the higher subbands, unless of course the code rate is increased. This tradeoff is further discussed in the following chapter where the performance of the system is evaluated.

4.2.1 Coding System

A complete system block diagram which includes the calculation of the masking threshold and the subband rate allocation is shown in Figure 4.5. Here, the steps involved in the encoding and decoding processes are shown more clearly. The input signal, $x(n)$, is processed in blocks of 224 samples in length. The motivation for block processing is explained in the following chapter. Encoding consists of decomposing each block into subband components and quantizing these components with a uniform scalar quantizer whose step size is determined through the assignment of subband rates. Since these rates depend directly on the subband's signal-to-mask ratio, the masking threshold must be computed prior to rate allocation and quantization. A 1024-point FFT is employed to determine the input's spectral components which are used in the computation of the masking threshold. The subband SMR values are computed next, and following is the allocation of rate among subbands. The subband quantizer step sizes are then determined from a look-up procedure which matches each nonzero subband rate with the closest predetermined rate based on the unit variance subband rate-distortion model. The quantizer step size corresponding to each matched, nonzero rate is chosen. Subbands receiving an allocated rate of zero are not transmitted to the decoder. Quantization and subsequent arithmetic coding of all subband components is then performed, along with the coding of side information consisting of subband variance values and table indices corresponding

to the subband rates. The decoding process is simply the inverse of the encoding process, however it is much less complex since it is not necessary to recompute the masking threshold and carry out an additional rate allocation. The decoding steps consist of arithmetically decoding all quantized subband samples and side information, performing the inverse quantization to form the subband sample values before quantization, and synthesizing the output, $\hat{x}(n)$, from the subband components. As shown in the following chapter, this output represents an accurate approximation to the input.

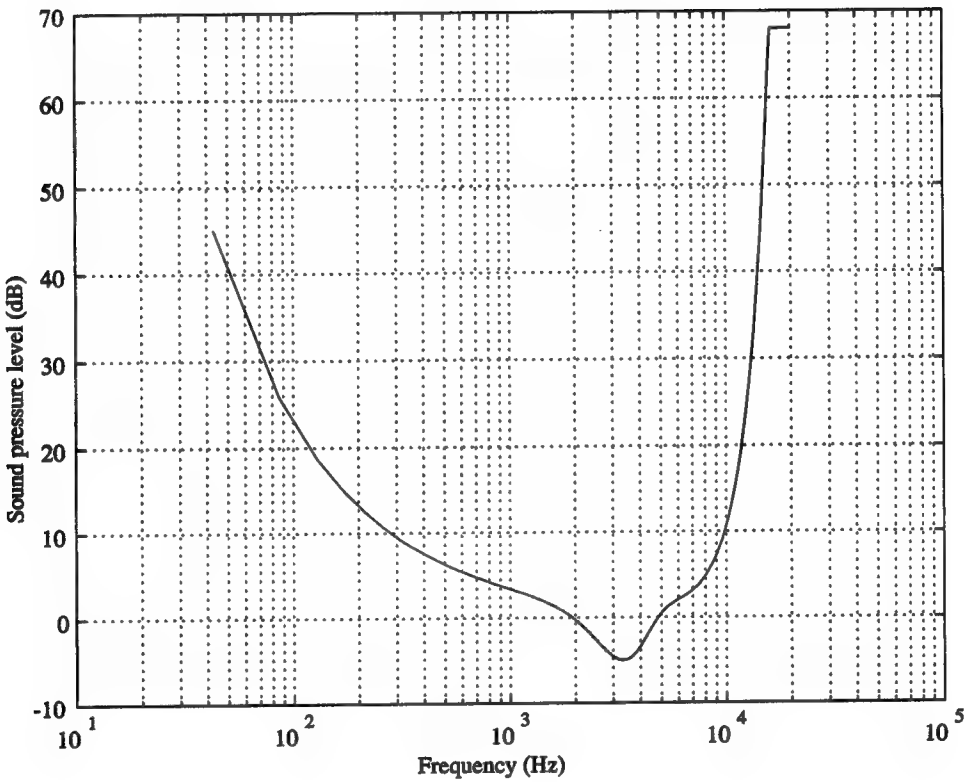

Figure 4.1: Threshold in quiet

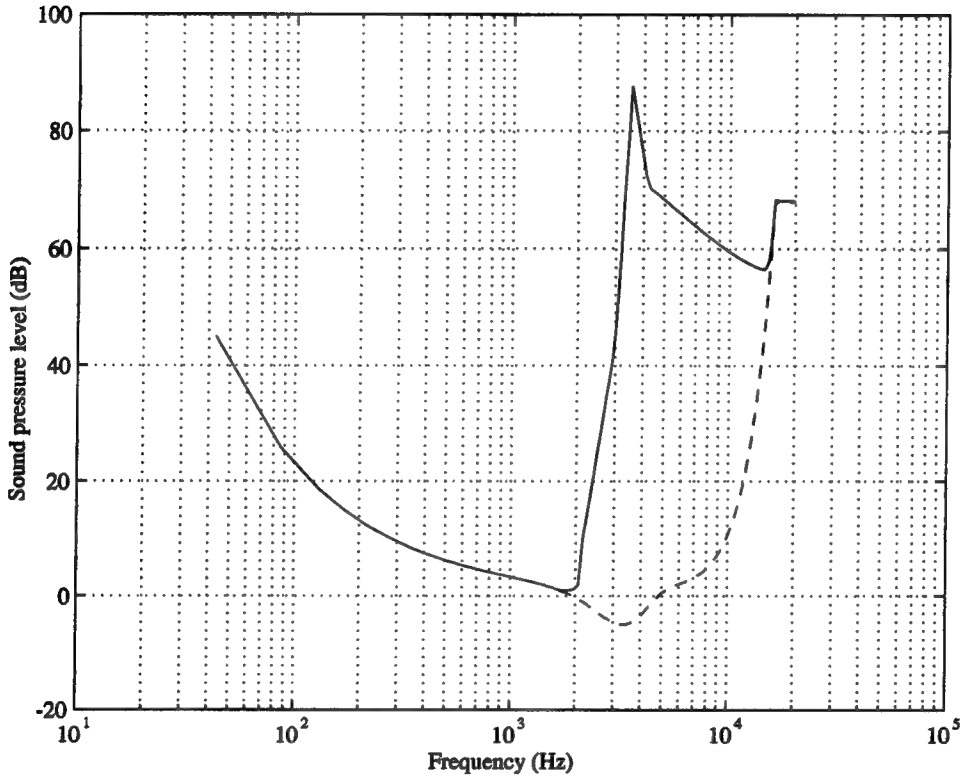

Figure 4.2: Masking threshold for a pure tone masker at 3.5 kHz

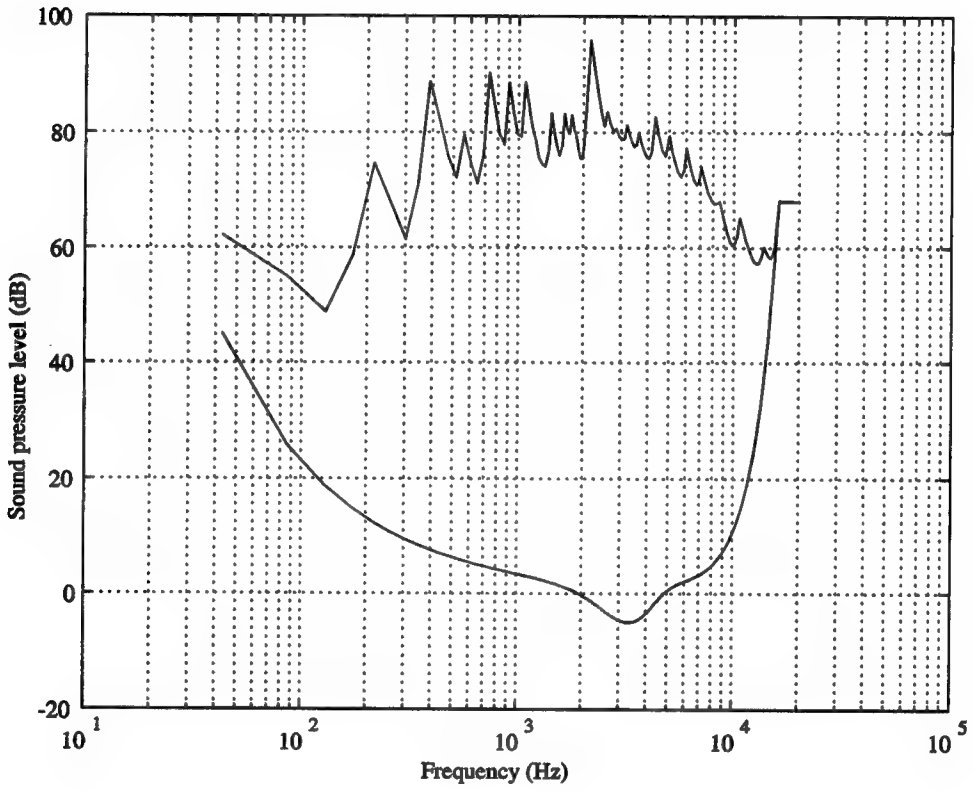

Figure 4.3: Masking threshold for a 23.2 ms fragment of classical music (upper curve), and threshold in quiet (lower curve)

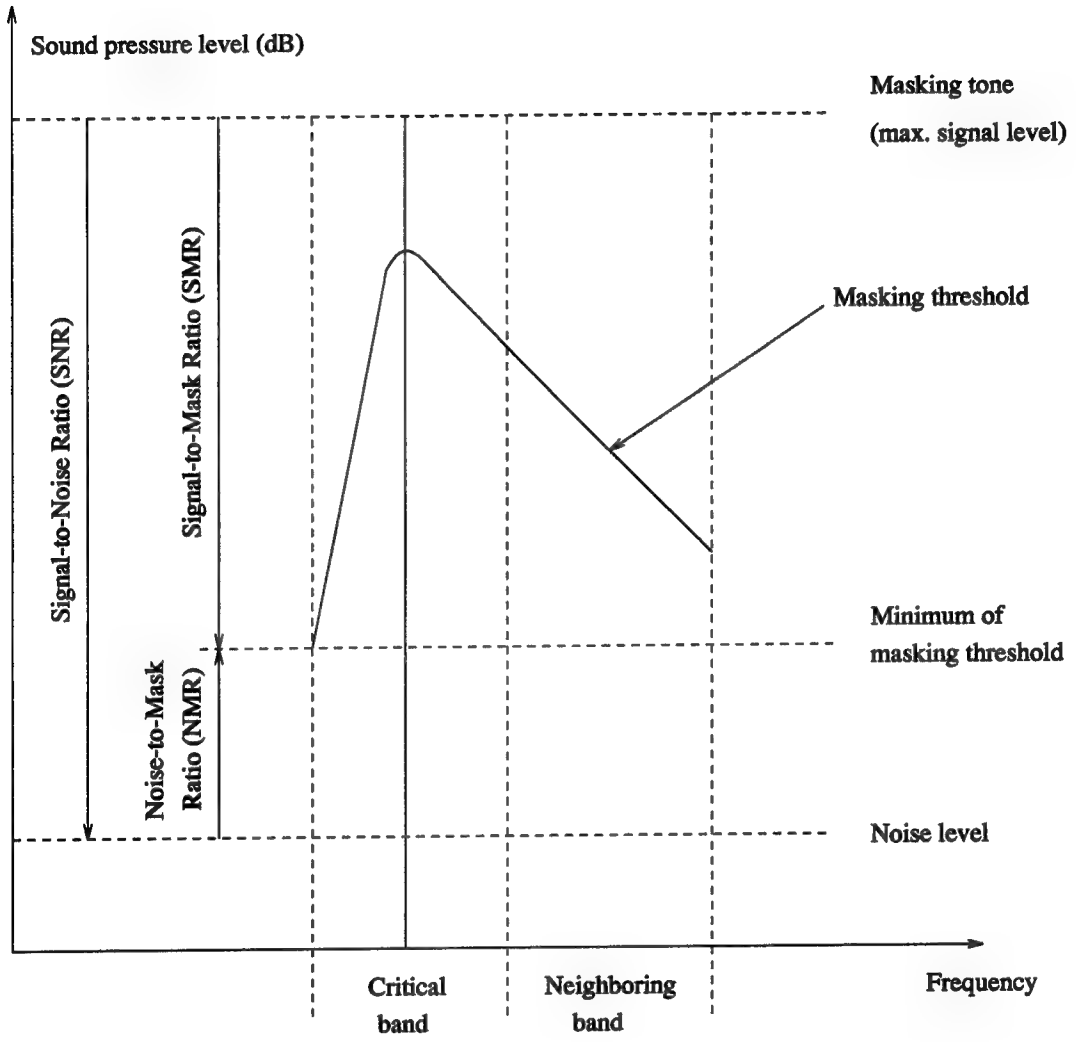

Figure 4.4: Illustration of signal-to-mask ratio (SMR)

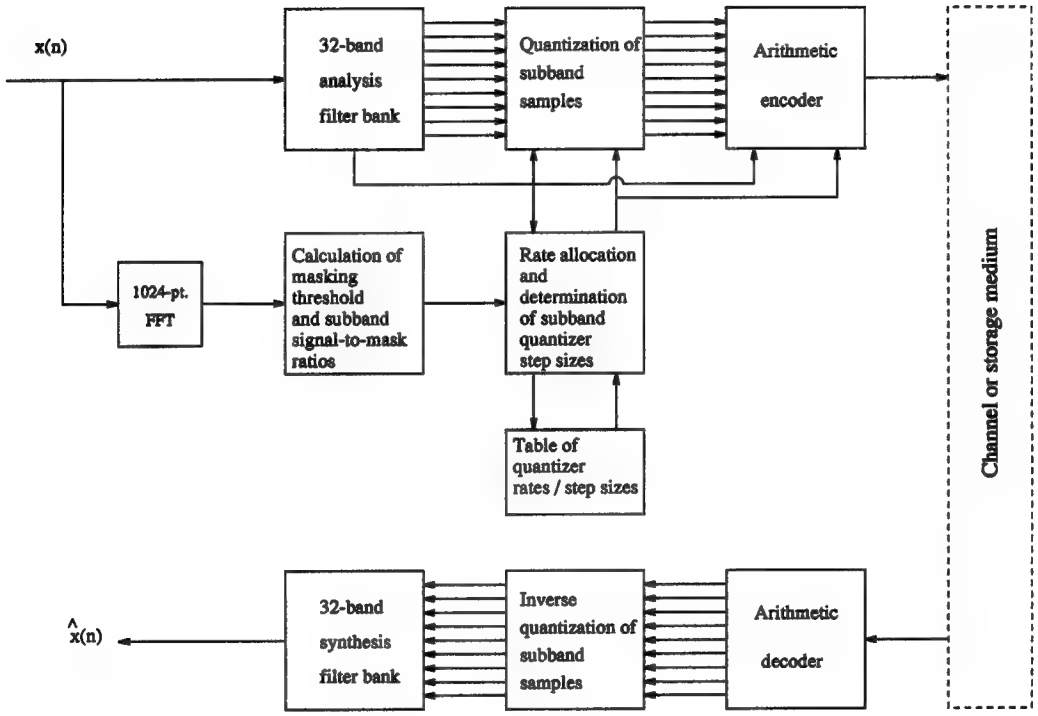

Figure 4.5: Complete system block diagram

CHAPTER 5
CODING SIMULATIONS AND PERFORMANCE RESULTS

The performance of the coding system was evaluated by implementing simulations on monophonic digital audio segments[1] taken from a compact disc. There were six audio segments used in the simulations with content ranging from classical to hard rock. System performance was based on both objective and subjective criteria. Measures such as mean-squared error (MSE) and signal-to-noise ratio (SNR) were used as an evaluation of the objective quality of the coding system. Also included was a measure of compression efficiency. This was determined by comparing the attainable bit rate to the theoretical entropy of the quantized audio source. The subjective quality of the system was determined using informal listening tests.

5.1 Objective Performance

The audio test segments used in the simulations were taken from a compact disc. These segments were sampled at 44.1 kHz and the samples were represented by 16-bit 2's complement integers. Conversion to floating point format was done before the samples were processed. The audio segments are listed according to content and length. A six letter string is used to distinguish one segment from another. This information is given in Table 5.1. For each of the coding simulations, the code rate, R, was fixed at 2 bits/sample. The input audio data was processed in blocks which were equal to 224 samples in length, or approximately 5.08 ms. Block processing was necessary in order to track changes in signal statistics and masking threshold characteristics. Similarity between the original and reconstructed audio segments

[1]The term "segment" is used to indicate a digital sequence representing a portion of an entire song or movement.

was measured by computing the MSE and SNR upon completion of each simulation. The MSE is defined as

$$MSE = E[(x(n) - \hat{x}(n))^2] \tag{5.1}$$

where $x(n)$ and $\hat{x}(n)$ represent the original and reconstructed audio sequences, respectively. The SNR is given by

$$SNR = 10\log_{10}(\frac{\sigma_x^2}{MSE}) \tag{5.2}$$

where σ_x^2 represents the signal power associated with $x(n)$. Table 5.2 contains a summary of the MSE and SNR calculations for each of the audio segments. This data shows fairly consistent system performance across all types of audio segments, which is indicative of the robustness of the system using the subband rate-distortion model defined in Chapter 3. Waveform plots of portions of the original and reconstructed audio sequences are provided to accompany the results in Table 5.2. These plots are shown in Figures $5.1 - 5.4$. It can be seen from the plots that the displayed portions of the reconstructed sequences are nearly identical to the corresponding portions of the original sequences, which demonstrates very accurate reconstruction capabilities.

Compression efficiency was evaluated by comparing the fixed code rate, R, to both the actual code rate obtained after encoding, and the theoretical entropy. Comparisons are based on the rates associated with the subband samples. Side information, which consists of rate allocation information and subband variances, is neglected for the moment. The fixed code rate is given in Equation 3.15. This quantity represents the mean of the assigned subband rates. The subband rates, in this case, are the entropies of the subband quantizers. For each of the simulations, R was set to 2 bits/sample, as mentioned above. The actual code rate was determined by computing the ratio of the number of bits needed to represent the subband

samples to the total number of subband samples sent to the decoder. This quantity will be denoted by R_A. The entropy was computed using the discrete version of Equation 3.13, which is given as

$$H = -\sum_i P_i \log_2(P_i) \tag{5.3}$$

where P_i represents the probability of the index of a particular quantizer level. Table 5.3 summarizes the resulting rates after coding each of the audio segments. It can be seen that the rates agree moderately well, especially for the coding of the classical audio segments. Deviation from the fixed rate may be attributed to the inaccuracy of the model in describing the true rate-distortion characteristics of the quantization of the subband signals. The average difference between the fixed and actual code rates was 0.102 bits/sample with a maximum difference of 0.170 bits/sample. Comparison between the entropy and the actual code rate gave an average difference of 0.019 bits/sample with a maximum difference of approximately 0.020 bits/sample. The total code rate was determined by including necessary side information that had to be made available to the decoder. This information consisted of table indices corresponding to the subband rates, and also subband variances. The average increase in code rate due to the side information was approximately 0.959 bits/sample. The actual rate, R_A, in addition to this increase forms the total code rate. Values for the total code rate are shown in Table 5.4, along with the transmission rate, which is defined as the product of the total code rate and the sampling frequency. The transmission rate, given in units of kbits/second, was computed for a sampling frequency of 44.1 kHz. Without compression, the transmission rate for CD-quality audio is approximately 706 kbits/second. Based on the simulations performed here, the average code rate was approximately 3.06 bits/sample, which corresponded to an average transmission rate of 135 kbits/second. At this rate, the overall reduction

factor was 5.23.

5.2 Subjective Performance

Listening tests were conducted to evaluate the perceptual quality of the coded audio segments. The outcomes of the tests were quantified using the method of mean opinion scoring. This method required subjects to classify the audio segments using a 5-point grading scale. Two 5-point grading scales are currently in use. One is used for signal quality, while the other is used for signal impairment. The impairment scale was used to grade the audio segments in this work. Each impairment level has an associated number score and label, which describe the differences between the original and coded audio segments, or equivalently, the noise content of the coded audio segment. These number scores and corresponding labels are shown in Table 5.5. Subjects were provided with three headphone presentations of each of the audio segments. The first of three was always the original, while the remaining two were the original and coded audio segments presented in an unknown order. It was the task of the subject to decide which of the two was the original and to grade the remaining one based on the amount of signal degradation present, if any. The audio segments were graded over ten trials, where each trial required grading of all six audio segments. The average of the scores taken over the ten trials was computed for each segment. This average is otherwise known as the mean opinion score (MOS) [31]. Resulting MOS values are shown in Table 5.6. With the exception of the scores for the classical audio segments, these values indicate very good subjective quality. Since the classical audio segments consisted of sounds which were very pure relative to the content of the other segments, it was fairly easy for subjects to make the distinction between the original and the coded segments. Noticeable sound differences, however, could have been diminished or eliminated

by increasing the code rate. As a figure of merit for the subjective quality of the system, the mean MOS value over all audio segments was computed. This value was equal to 4.4, which would indicate perceptible but tolerable differences in sound quality between the original and coded audio segments.

Audio Segment	Content	Length (seconds)
MAHHOR	classical	5.44
MOZSTR	classical	8.80
STGILF	soft rock	5.76
FWMDRM	soft rock	6.88
PJMANL	hard rock	8.96
STPWGN	hard rock	8.32

Table 5.1: Audio test segments

Audio Segment	MSE ($\times 10^{-5}$)	SNR (dB)
MAHHOR	1.2008	29.9088
MOZSTR	0.6194	32.1598
STGILF	2.5525	29.7843
FWMDRM	1.7050	28.0540
PJMANL	7.5012	29.0491
STPWGN	5.9240	28.7387

Table 5.2: Mean-squared error and signal-to-noise ratio results

Audio Segment	Fixed Rate, R (bits/sample)	Actual Rate, R_A (bits/sample)	Entropy, H (bits/sample)
MAHHOR	2	2.0427	2.0235
MOZSTR	2	2.0094	1.9903
STGILF	2	2.1169	2.0971
FWMDRM	2	2.1529	2.1343
PJMANL	2	2.1227	2.1040
STPWGN	2	2.1702	2.1511

Table 5.3: Comparison among fixed and attainable code rate, and quantized source entropy

Audio Segment	Total Rate (bits/sample)	Transmission Rate (kbits/second)
MAHHOR	3.0190	133
MOZSTR	2.9748	131
STGILF	3.0771	136
FWMDRM	3.1060	137
PJMANL	3.0739	136
STPWGN	3.1202	138

Table 5.4: Code rate including overhead and corresponding transmission rate at a sampling frequency of 44.1 kHz

Number Scores	Impairment Scale
5	Imperceptible
4	Perceptible but not Annoying
3	Slightly Annoying
2	Annoying
1	Very Annoying

Table 5.5: Five-point adjectival grading scale for signal impairment

Audio Segment	MOS Value
MAHHOR	3.6
MOZSTR	3.3
STGILF	5.0
FWMDRM	4.9
PJMANL	4.8
STPWGN	4.5

Table 5.6: Mean opinion score (MOS) results

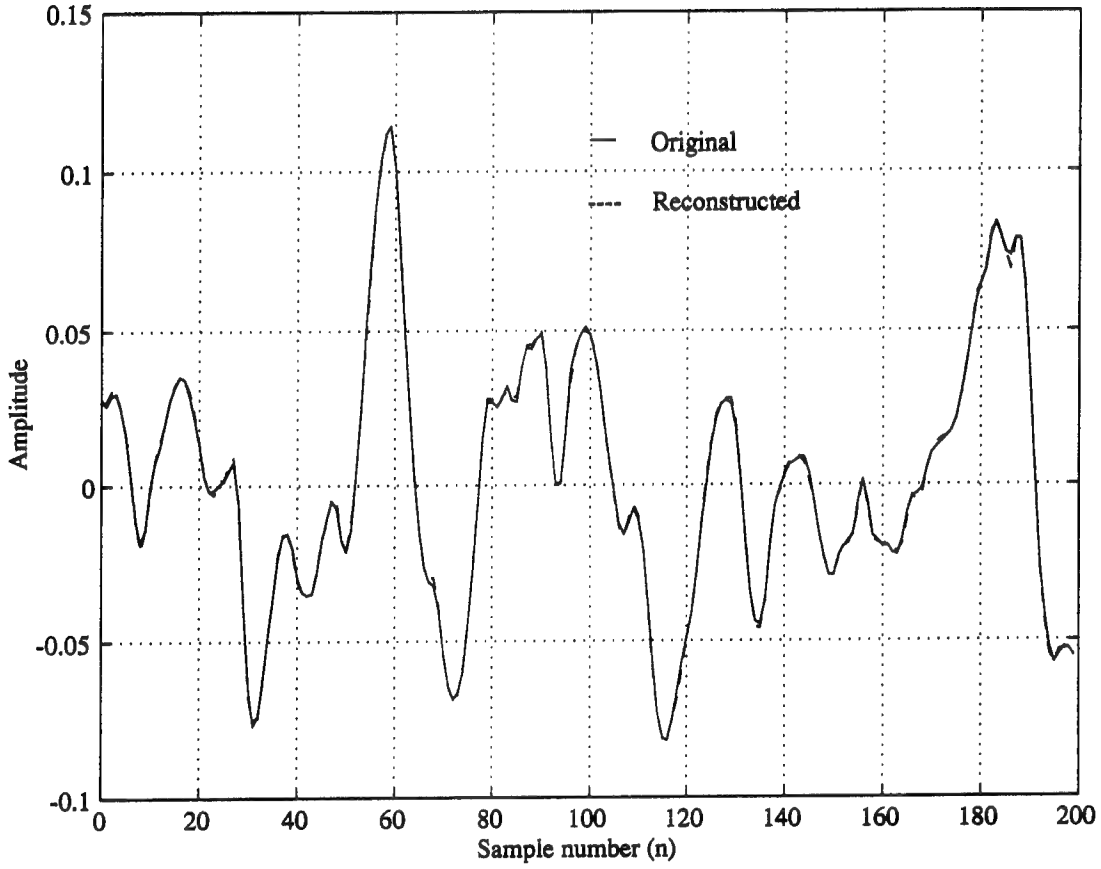

Figure 5.1: Portion of original, $x(n)$, and reconstructed, $\hat{x}(n)$, "MAH-HOR" audio segment

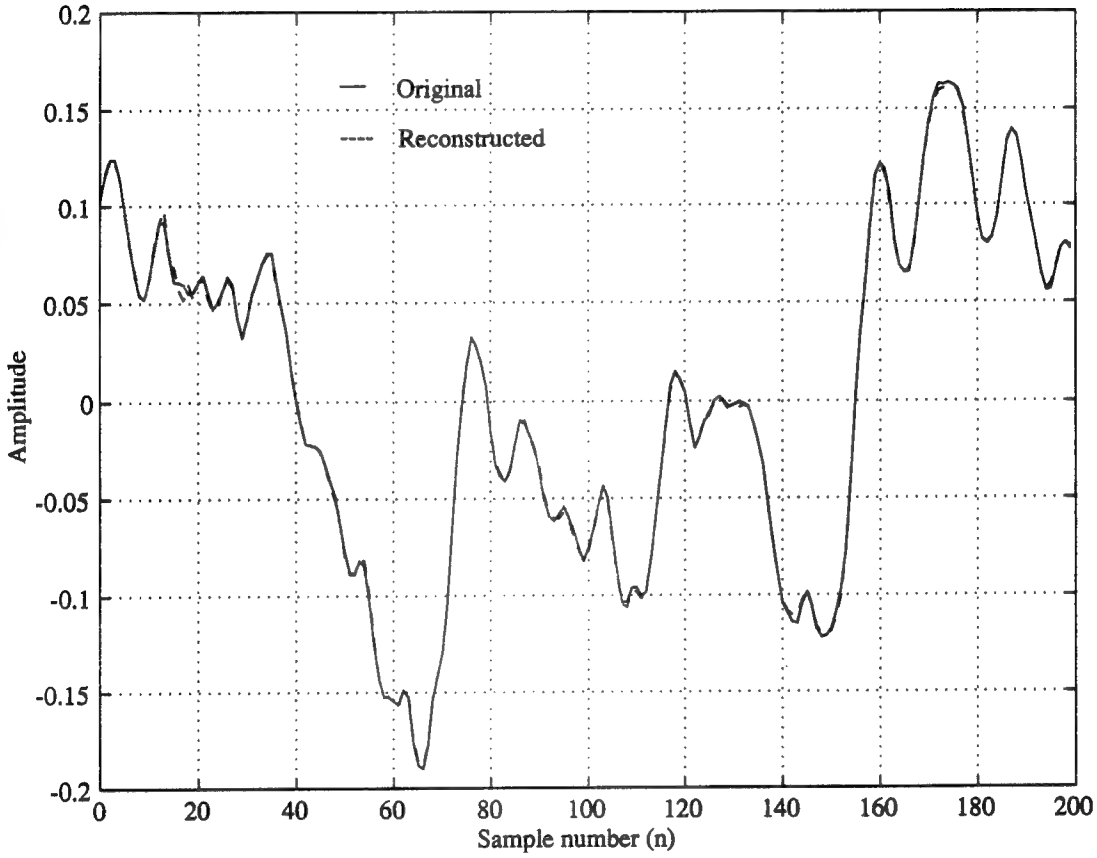

Figure 5.2: Portion of original, $x(n)$, and reconstructed, $\hat{x}(n)$, "MOZSTR" audio segment

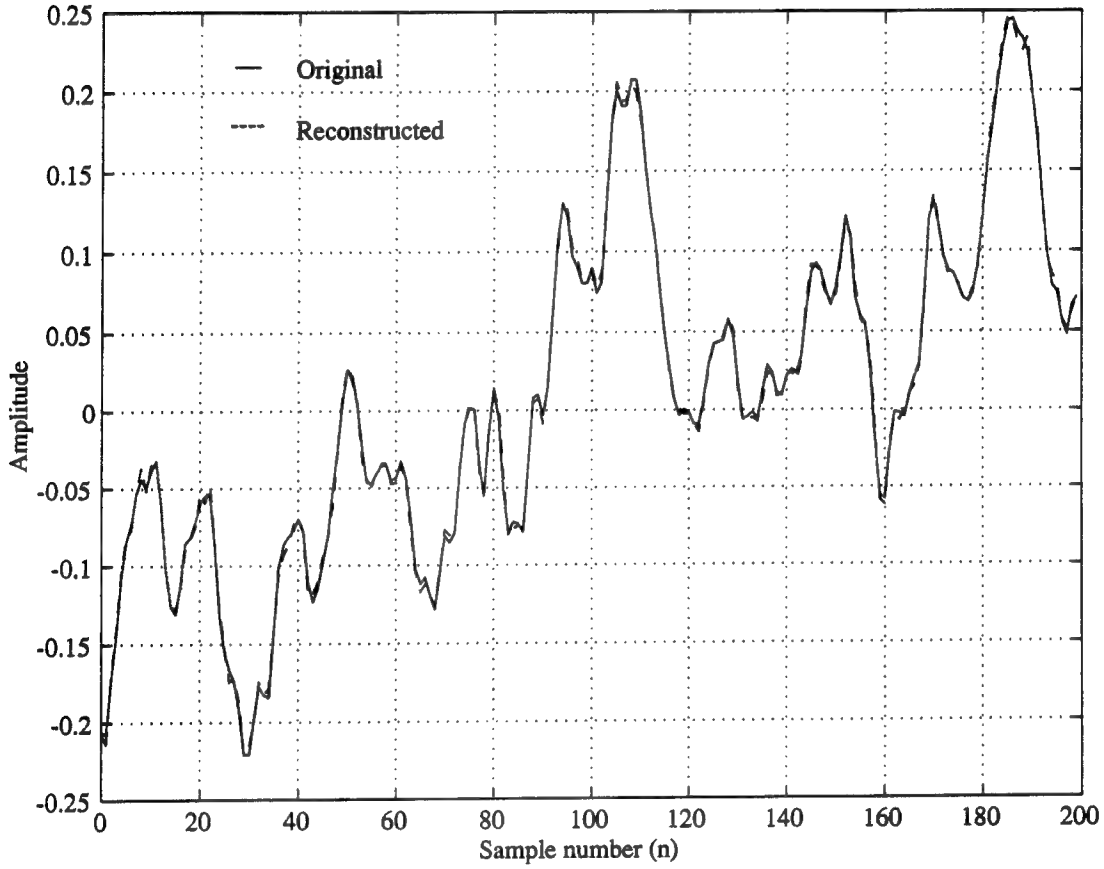

Figure 5.3: Portion of original, $x(n)$, and reconstructed, $\hat{x}(n)$, "STGILF" audio segment

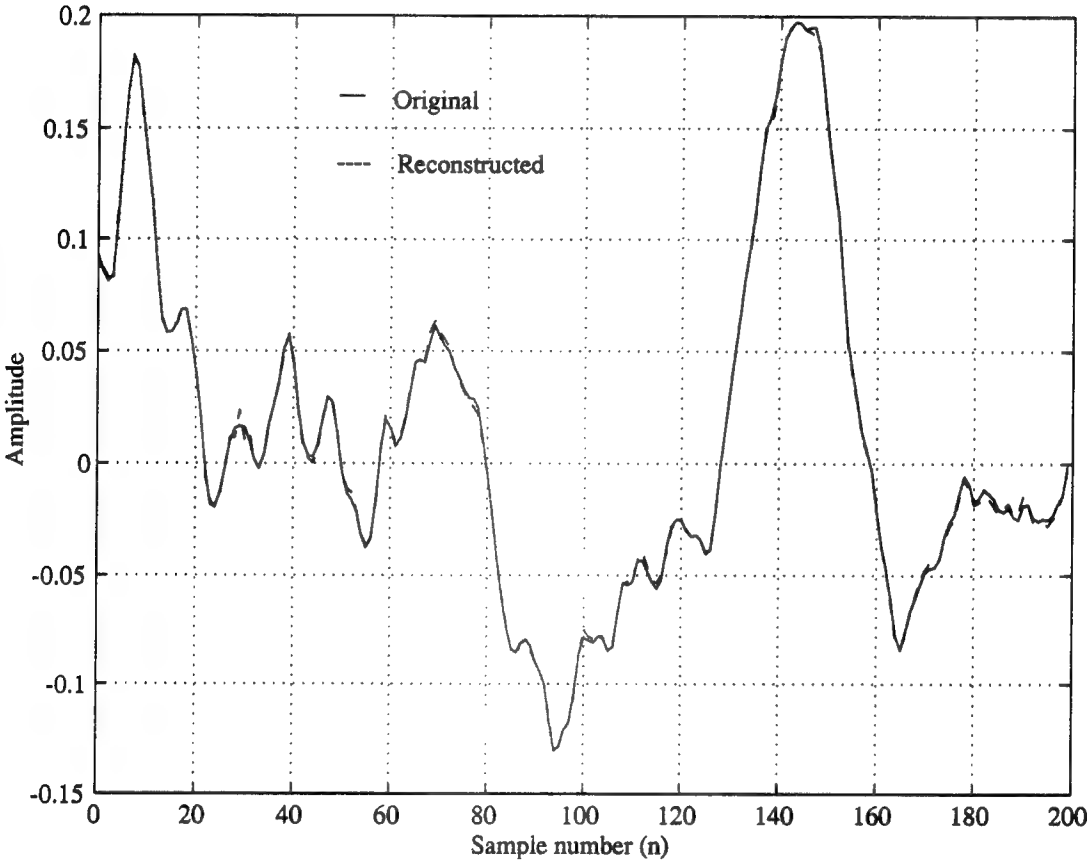

Figure 5.4: Portion of original, $x(n)$, and reconstructed, $\hat{x}(n)$, "FWM-DRM" audio segment

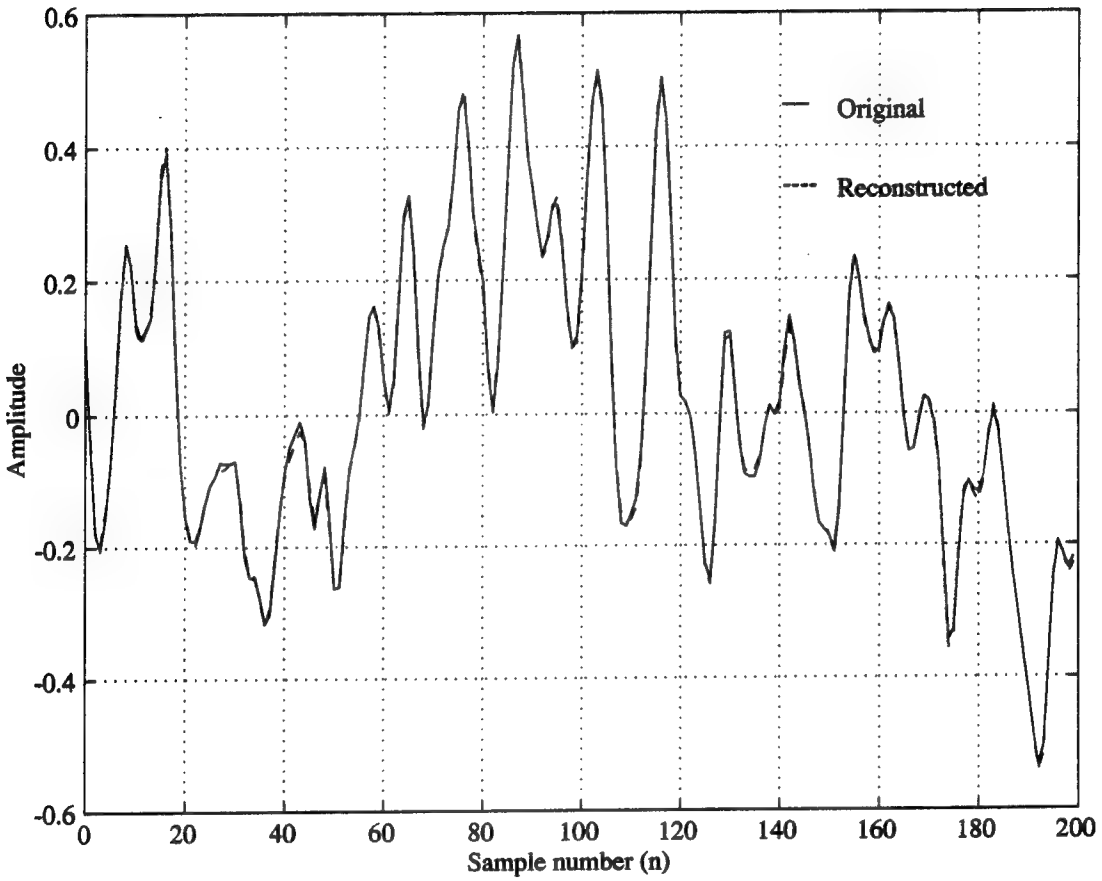

Figure 5.5: Portion of original, $x(n)$, and reconstructed, $\hat{x}(n)$, "PJ-MANL" audio segment

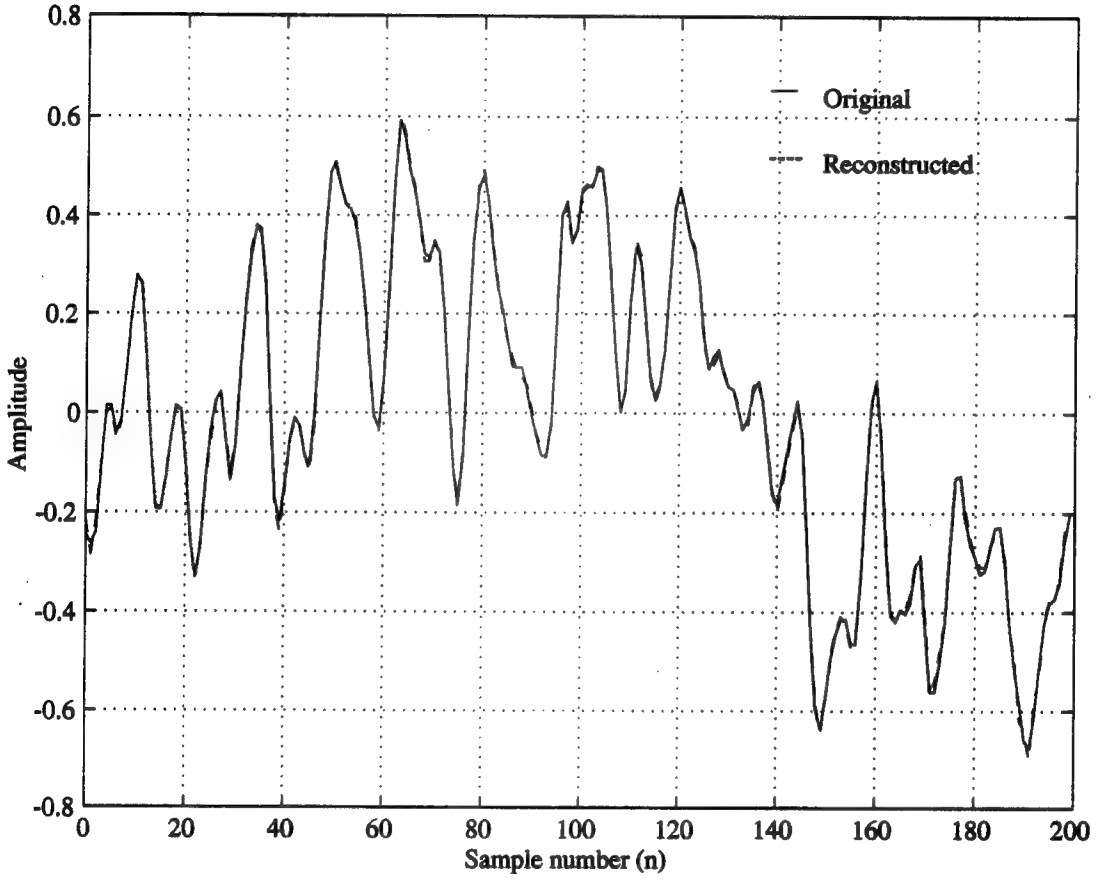

Figure 5.6: Portion of original, $x(n)$, and reconstructed, $\hat{x}(n)$, "STP-WGN" audio segment

CHAPTER 6
CONCLUSION AND DISCUSSION

A system designed for high quality coding of wideband audio has been presented. This system employed a 32-band single-sideband modulated filter bank to perform subband analysis and synthesis operations. Encoding and decoding of the subbands was accomplished using entropy-constrained scalar quantization and subsequent entropy coding. The subband quantizers contained uniform decision thresholds and output levels. The number of output levels was 255. Each subband quantizer was constructed using the subband variance and a uniform step size that corresponded to the required quantizer entropy given by the subband rate allocation. The subband quantizer indices, along with side information consisting of subband variances and indices used to identify the subband rates, were entropy coded. The particular form of entropy coding used in this system was arithmetic coding. An adaptive probability model was used in the arithmetic encoding and decoding algorithms. The subband rate allocation procedure that was applied relied on analytic models to describe the rate-distortion characteristics of the subband quantizers. This method required minimal computations and resulted in fast convergence for a desired code rate. The use of the masking threshold was included in the rate allocation procedure as a weighting factor on the subband distortion terms. This allowed for the placement of greater emphasis on the subbands which required larger rates, or equivalently, larger signal-to-noise ratios. The performance of the system was primarily assessed using comparisons among the code rates and the entropy, as well as mean opinion score results. At a fixed code rate of 2 bits/sample, the actual code rate deviated by an average of 0.102 bits/sample. The average difference between the entropy

and the actual code rate was only 0.019 bits/sample. The average value of the total code rate, or the actual rate plus side information, was 3.06 bits/sample, which corresponded to an average transmission rate of 135 kbits/second. At this rate, the system yielded a mean MOS value of 4.4, which was a reasonably good result considering there were just six audio segments used in the subjective evaluations.

The system presented in this thesis suggested a unique approach to the coding of wideband audio signals. In the absence of the overhead, or side information, code rate, an attempt was made to achieve an actual code rate equivalent to the entropy of the quantized source samples. Unlike previous audio coding systems, the approach taken here employed the technique of entropy-constrained scalar quantization, and also utilized analytic models to describe the rate-distortion characteristics of the subband quantizers. The models were derived from the entropy-distortion characteristic for the uniform quantization of a unit variance Laplacian signal. The assumption of the Laplacian pdf, which was used to describe the subband signals, provided the basis for a simple and sufficiently accurate rate-distortion model which could be easily incorporated into the subband rate allocation procedure. Entropy-constrained quantization allowed for non-integer subband rates and did not require the number of quantization levels to vary in order to achieve various degrees of quantizer performance across the subbands. Instead, quantizer performance was based on the assignment of different quantizer step sizes to each subband. These step sizes were determined through the assigned rates resulting from the subband rate allocation procedure. One of the benefits of this rate allocation and quantization scheme was improved code rate performance over quantization without an entropy constraint. This result is due to the fact that the quantizer output entropy is always less than the base two logarithm of the number of quantizer output levels, unless of course the levels are equally probable. A second benefit is that this scheme was

computationally efficient, since the rate allocation procedure was known to converge quickly and the quantizer step size values which corresponded to the subband rates could be found by simply performing a table look-up.

Simulations have shown that for all of the audio segments encoded, an actual code rate fairly close to the entropy could be achieved with minimal or imperceptible differences in signal quality between the original and coded segments. This is an interesting result since previous systems have focused on achieving code rates bound by the "perceptual entropy." This quantity has been used to define the minimal code rate needed to maintain transparent differences between the original and coded segments. It should be noted that in order to have achieved the quantizer output entropy, the masking constraints were satisfied for only a portion of the audio frequency band. In some cases, this was not perceptibly tolerable. Another disadvantage that resulted was the increase in code rate due to the rate of the side information. The average increase was 0.959 bits/sample, which was quite substantial and clearly undesirable. Contributing most to this increase were the subband variances which were used by the decoder to scale the subband quantizers. A reasonable solution would be to quantize these variances in order to reduce the number of bits used to represent them. The code rate increase due to side information caused the average of the total code rates to be equal to approximately 3.06 bits/sample when the fixed code rate was set to 2 bits/sample. The transmission rate corresponding to this total code rate was 135 kbits/second per monophonic channel, and the mean MOS value was 4.4. These performance results fall in between the results of the Layer I and Layer II MPEG audio coders. The sampling rate used in the MPEG evaluations was 48 kHz. Evaluations of the Layer I MPEG audio coder resulted in a mean MOS value (over 10 test items) of 4.7 at a rate of 192 kbits/second per monophonic channel, while Layer II MPEG evaluations resulted

in a mean MOS value of 4.8 at a rate of 128 kbits/second per monophonic channel [11]. The equivalent code rate for Layer I MPEG was 4 bits/sample, while for Layer II it was 2.67 bits/sample.

The system presented in this thesis, while comparable in its rate reduction ability, gives lower subjective performance than either the Layer I or Layer II MPEG coders. The subjective performance can be improved, however, by simply increasing the code rate. In order to satisfy both performance goals, the code rate increase caused by side information must be reduced. The system will then be able to provide the desired reductions in rate and maintain the high level of signal quality that is so essential in wideband audio coding systems.

BIBLIOGRAPHY

[1] C. E. Shannon, "A mathmetical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July 1948.

[2] T. Berger, *Rate Distortion Theory.* Englewood Cliffs, NJ: Prentice Hall, 1971.

[3] R. G. Gallager, *Information Theory and Reliable Communication.* New York: McGraw Hill, 1965.

[4] L. D. Davisson, "Rate-distortion theory and application," *IEEE Proc.*, pp. 800–808, July 1972.

[5] R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital coding of speech in subbands," *Bell Syst. Tech. J.*, vol. 55, pp. 1069–1085, October 1976.

[6] R. E. Crochiere, "On the design of subband coders for low-bit-rate speech communication," *Bell Syst. Tech. J.*, vol. 56, pp. 747–770, May-June 1977.

[7] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," *IEEE Trans. Acoust., Speech, and Signal Proces.*, pp. 299–309, August 1977.

[8] J. D. Johnston and D. J. Goodman, "Digital transmission of commentary-grade (7 kHz) audio at 56 or 64 kb/s," in *Proc. ICASSP*, pp. 442–444, 1979.

[9] R. V. Cox, "A comparison of three speech coders to be implemented on the digital signal processor," *Bell Syst. Tech. J.*, vol. 60, pp. 1411–1421, September 1981.

[10] R. E. Crochiere, "Subband coding," *Bell Syst. Tech. J.*, vol. 60, pp. 1633–1653, September 1981.

[11] P. Noll, "Wideband speech and audio coding," *IEEE Commun. Mag.*, pp. 34–44, November 1993.

[12] K. Brandenburg, "OCF - A new coding algorithm for high quality sound signals," in *Proc. ICASSP*, pp. 141–144, 1987.

[13] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, pp. 314–323, February 1988.

[14] K. Brandenburg and G. Stoll, "The ISO/MPEG-Audio Codec: A generic standard for coding of high quality digital audio," in *92nd AES Convention*, (Vienna), March 1992. Preprint No. 3336.

[15] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1983.

[16] J. H. Rothweiler, "Polyphase quadrature filters - A new subband coding technique," in *Proc. ICASSP*, pp. 1280–1283, 1983.

[17] P. L. Chu, "Quadrature mirror filter design for an arbitrary number of equal bandwidth channels," *IEEE Trans. Acoust., Speech, and Signal Proces.*, pp. 203–218, February 1985.

[18] P. P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions, and perfect-reconstruction techniques," *IEEE ASSP Mag.*, pp. 4–20, July 1987.

[19] K. Nayebi, T. P. Barnwell, III, and M. J. T. Smith, "On the design of FIR analysis-synthesis filter banks with high computational efficiency," *IEEE Trans. Signal Proces.*, pp. 825–834, April 1994.

[20] ISO/IECJTC1/SC2/WG11 MPEG 91/0.., *Coding of Moving Pictures and*

Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s, November 1991.

[21] W. A. Pearlman, "Chapter 2 : Information Theory and Image Coding," Class notes, 1991.

[22] T. J. Goblick, Jr. and J. L. Holsinger, "Analog source digitization: A comparison of theory and practice," *IEEE Trans. Info. Theory*, pp. 323–326, April 1967.

[23] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Info. Theory*, pp. 676–683, September 1968.

[24] T. Berger, "Optimum quantizers and permutation codes," *IEEE Trans. Info. Theory*, pp. 759–765, November 1972.

[25] G. G. Langdon, Jr., "An introduction to arithmetic coding," *IBM J. Res. Develop.*, vol. 28, pp. 135–149, March 1984.

[26] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. of the ACM*, vol. 30, pp. 520–540, June 1987.

[27] P. H. Westerink, J. Biemond, and D. E. Boekee, "An optimal bit allocation algorithm for subband coding," in *Proc. ICASSP*, pp. 757–760, 1988.

[28] E. A. Riskin, "Optimal bit allocation via the generalized BFOS algorithm," *IEEE Trans. Info. Theory*, pp. 400–402, March 1991.

[29] R. N. J. Veldhuis, "Bit rates in audio source coding," *IEEE J. Select. Areas Commun.*, pp. 86–96, January 1992.

[30] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models.* Berlin, Germany: Springer-Verlag, 1990.

[31] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video.* Englewood Cliffs, NJ: Prentice Hall, 1984.

*U.S. GOVERNMENT PRINTING OFFICE: 1995-610-126-50215

Rome Laboratory

Customer Satisfaction Survey

RL-TR-_____

Please complete this survey, and mail to RL/IMPS, 26 Electronic Pky, Griffiss AFB NY 13441-4514. Your assessment and feedback regarding this technical report will allow Rome Laboratory to have a vehicle to continuously improve our methods of research, publication, and customer satisfaction. Your assistance is greatly appreciated.
Thank You

_____

_____

Organization Name:_____(Optional)

Organization POC: _____(Optional)

Address:_____

1. On a scale of 1 to 5 how would you rate the technology developed under this research?

5-Extremely Useful 1-Not Useful/Wasteful

Rating_____

Please use the space below to comment on your rating. Please suggest improvements. Use the back of this sheet if necessary.

2. Do any specific areas of the report stand out as exceptional?

Yes___ No_____

If yes, please identify the area(s), and comment on what aspects make them "stand out."

3. Do any specific areas of the report stand out as inferior?

Yes___ No___

If yes, please identify the area(s), and comment on what aspects make them "stand out."

4. Please utilize the space below to comment on any other aspects of the report. Comments on both technical content and reporting format are desired.

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.